SCIENCE AND TECHNOLOGY BRIEFINGS
PARLIAMENTARY OFFICE FOR SCIENTIFIC AND TECHNOLOGICAL ASSESSMENT

Briefing 36

January 2023

The Data Explosion: Preventing Flooding

N° 768 ASSEMBLÉE NATIONALE – N° 291 SÉNAT

Source: sekulicn

Summary

- *Today's data explosion is an undeniable reality, the full extent of which requires further examination. Even though there have been other data deluges in the past, the scale of this one is unprecedented.*
- *Quite apart from its environmental impact, the current explosion poses, directly or indirectly, various challenges for our societies, including the attention economy, filter bubbles, information overload and new inequalities.*
- *To prevent the risk of being swamped by this tidal wave, technological solutions exist, but a cultural battle is needed to promote initiatives to minimise the obsessive accumulation of data, which has turned into a "digital Diogenes syndrome".*

Ludovic Haye, Senator

A recurring issue with a long history

Collecting, accumulating and processing data were human activities long before automatic means emerged to perform such tasks. Institutions such as the Church and the State were responsible for this work, which dates as far back as the earliest urban civilisations of Mesopotamia and Egypt.

As historian Robert Darnton points out, *"every age was an age of information, each in its own way"*.[1]

Data explosion is not then a new phenomenon, and earlier episodes may have created fears that have since been forgotten [2].

Nevertheless, previous data deluges seem quite moderate compared with today's era of digital civilization and Big Data, defined by the "3 Vs" of data volume, velocity and variety.

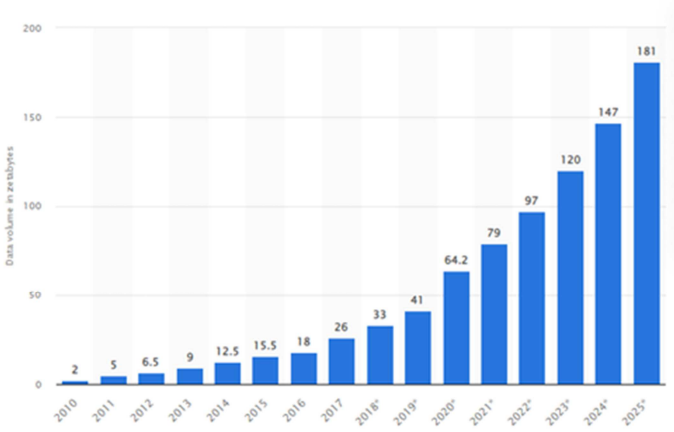
Gauging the extent of the data explosion

The volume of data created is growing exponentially, rising globally from two zettabytes in 2010 to 18 zettabytes in 2016, and then, according to forecasts, to 64 zettabytes in 2020 and 181 zettabytes in 2025.[3]

This highlights the challenge of data storage, an issue already addressed in a separate Science and Technology Briefing[4] and which will not be examined here.

The human and technological capabilities required to process and sort this data will have difficulty keeping up with its rate of growth, due to very high costs and the fact that resources are concentrated around a limited number of stakeholders.

Growth in data volume (in zettabytes)



Source: Statista

According to Daniel Rosenberg, data is "a category of facts and principles that were, by agreement, beyond argument":[5] data is therefore variable in content, constructed by social and historical consensus. For example, letters were rarely saved in the past and were therefore not data, but today's emails are data as they are retained for scanning by existing and future algorithms.

Furthermore, whereas in the past the State, together with the Church, was the main collector of data, today it is private companies that collect data from multiple sources, including the internet, physical sensors, cameras, satellites, navigation

systems, industrial processes, health, scientific activities and genomics. These developments highlight the unique nature of the current data explosion.

There are several typologies designed to categorise data according to various criteria, including their nature, structure and use, in order to understand, visualise and, if necessary, process the data.(6) The fact that the amount of inaccurate data is rapidly rising also makes it more difficult to process all the available data.

New opportunities provided by Big Data

The hype surrounding Big Data analysis is primarily driven by the new opportunities it offers. While isolated data is of little use, when it is combined and processed, knowledge can be extracted and predictions made. While deductions traditionally use physical laws derived from prior empirical research of causes and determinants to make predictions through modelling, as in weather forecasting, digital technologies use Big Data through inductive approaches, analysing vast amounts of data to derive laws. There are two types of analysis: applying statistical models to Big Data, and machine learning, in which the machine builds the algorithmic predictive models that are central to current research in artificial intelligence or AI.(7)

Big Data offers many opportunities, particularly on an economic level, as the following metaphors illustrate: the new black gold, oil of the 21st-century and the big data revolution. Improvements in business efficiency and human resource management, the optimisation of customer relations through personalised and/or predictive marketing, and the monetisation of customer information are just some of the expected benefits for companies. There are many fields in which it can be used, including health, transport and science. In health, Big Data has been key to advances in epidemiology, diagnosis and treatment accuracy.(8) In transport, Big Data can be used to optimise use of different means of transport and to personalise the passenger experience.(9)

Data has also led to the emergence of new business models, introduced by companies such as Google, Facebook, Amazon and Twitter. These large corporations are using user data to operate their services, to generate revenue through targeted advertising,(10) often by selling this data as profiles auctioned through data brokers, and to predict user intentions – what people will buy or do – using algorithms. The data explosion has led many companies, even those outside the digital sector, to add data processing to their existing business activities(11) and to examine these activities in more details with a view to optimising them.

A worrying environmental and societal impact

Aside from its real or expected benefits, Big Data brings with it various challenges. Digital tools are far from being just "virtual"; their use has very real consequences. For example, their environmental impact is increasing by 6% per year: they were responsible for at least 3.5% of global greenhouse gas (GHG) emissions and 4.2% of global primary energy consumption in 2019.(12) A report by The Shift Project also points out that 44% of French people consider computers and the internet to be a threat to the environment.

An attention economy has been built around the use of Big Data, in which companies take advantage of data to capture users' attention even more effectively, expose them to even more advertising, and collect even more information in a circular fashion. This *"new capitalist model"*, which some describe as *"cognitive capitalism"*,(13) encourages digital companies firstly to trap users in filter bubbles,(14) which reinforce their points of view and beliefs and focus on subjects that seem to interest them, and secondly to spur consumers to constantly consume more as a consequence of this "influence industry", which in turn compounds the environmental problem.(15) This race for attention has also profoundly transformed journalism. The media's focus is now increasingly on speed of publication rather than quality of information.(16) From a political point of view, the shift towards mediocre, mass-produced information that seeks to attract rather than inform is a threat to democracy.

Democracy is also weakened by the mass surveillance that digital technology has made possible, compounded by the geopolitical issues inherent in Big Data, where the domination of American companies and their data centres is the hallmark of data governance in what is truly the age of surveillance capitalism.(17) Surveillance capitalism can be put to work for the State, as the *Prism* programme in the United States deftly demonstrated through its recovery, since 2007, of billions of pieces of personal data collected by companies from all over the world. Never one to be outdone, China has not hesitated to use TikTok for surveillance purposes, as the recent example of spying on foreign journalists illustrates.(18)

The Big Data era has also increased the inequalities associated with digital transformation. According to Eszter Hargittai, a second digital divide has emerged:(19) the first, linked to disparities in access to digital technologies, has narrowed, but the second, linked to the ability to adopt digital technologies (depending on social background and level of education), is widening. Furthermore, the use of Big Data can exacerbate existing inequalities because of biases in the algorithms whose results reproduce the biases of the data samples used. For example, COMPAS, the reoffending prediction software used by the US justice system, tends to overestimate the risk of reoffending among Black people.

The indirect link between data explosion and information overload

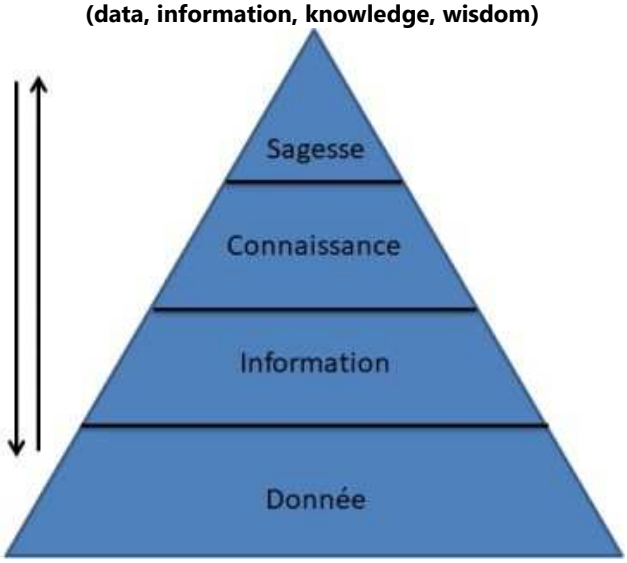
To understand the relationship between data explosion and information overload, it is first necessary to identify the links between data, information, knowledge and wisdom. These concepts are linked, but the boundaries between them are

blurred and need to be redefined. Data is an elementary description of a reality, which has little or no meaning when taken in isolation; it is an element or a set of objective elements about a fact. Information is a set of contextualised data from which meaning can be inferred. In understanding the meaning of the information and then integrating it into our system of knowledge and values, the information acquires the status of knowledge. Lastly, from knowledge, we acquire wisdom, or in other words, ethical behaviours.

Albert Einstein wrote that *"Knowledge is acquired through experience, everything else is just information"*[20] and T. S. Eliot asked *"Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?"*[21].

The relationship between these concepts can be illustrated through meteorology. The temperature measurements taken each day at a weather station are data. A curve showing the change in average temperature in a place over time is information. The fact that the Earth's temperature is rising as a result of human activity is knowledge. Adapting human activity according to its impact on global warming is wisdom. The diagram below illustrates these relationships.

DIKW Pyramid
(data, information, knowledge, wisdom)



Source: OPECST

Information volume (too much information to be processed by an individual or an organisation in a given time leading to information overload), cognitive overload (reaching the limits of the cognitive abilities of individuals to process a given volume of information) and lastly communication overload (new electronic means of communication – emails, online discussion forums, instant messaging, etc. – contribute significantly to information overload) are the three factors that combine to cause information overload, defined as a three-dimensional concept by Michel Kalika and other researchers.[22]

Information overload or infobesity

Information overload refers to the excess of information that a person cannot process or cope with without harming themselves. First coined in 1962, the term "information overload" was popularised in 1970 by the American futurologist Alvin Toffler. The concept of "infobesity" appeared in 1993 with David Shenk, who drew a symbolic parallel between information overload and obesity, in which the excess of information that stifles our intellectual processes is compared to the excess of fat that characterises obesity.

Difficulties linked to information overload and changes in our cognitive structures

As a report by France's Centre for Strategic Analysis and Directorate General for Labour points out, *"Information overload (or infobesity) is one of the biggest problems that organisations need to solve over the next ten years"*.[23] Information overload is a relatively new concept that describes the long-standing idea that human beings have a limited capacity to process information, both in terms of the quantity of information and the ability to retain information over a given period of time.[24] Of course, how people react to this information overload varies depending on the individual because, as Michel Kalika states, *"we are not all equal when it comes to information overload"*.[25]

There are several reasons for this, all of which are a consequence of the communication explosion, in terms of the quantity of messages received and the quantity of applications and communication channels used.[26] Information and communication technologies (ICT) have led to a significant increase in the quantity of both these aspects.

Information overload is also fuelled by the principles governing our social behaviour, including our addiction to communication, anxiety about failing to receive information (Fear of Missing Out or FOMO, a syndrome that can lead to depression and is particularly prevalent among young people, although it is now becoming more widespread), changes and instability specific to the world of work,[27] multitasking, the blurring of boundaries between work and private life, and the apparent free access to electronic communications.

According to Caroline Sauvajol-Rialland, emails in particular seem to be one of the main tools contributing to information and communication overload because *"the rate at which emails are exchanged has become too fast and uncertain. The exponential growth in volume makes it increasingly difficult for employees to manage and process their emails"*.[28]

This infobesity has many consequences, the main ones being: at the individual level, stress, anxiety, depression, reduced creativity, burnout; at the organisational level, reduced productivity, saturation, disorganisation; and at the societal level, wasted energy, a significant environmental (and especially carbon) footprint.

Another difficulty arises from the reasoning method used to process the data. If inductive reasoning is up-and-coming and deductive reasoning is, temporarily at least, declining, these two types of reasoning should coexist, otherwise Big Data risks bringing about the *"death of scientific theory"*.[29]

Attention should also be paid to the problems associated with the use of statistics: the ever growing volumes of data increase the opportunities to process them in ways that sometimes give the false impression of capturing the truth. In the words of two researchers, *"numbers are like people. If you torture them enough, you can make them say anything"*.[30] As Alain Desrosières wrote in a cautionary note, *"data are not given"*,[31] they are constructed and have no meaning in themselves, they only describe a part of what actually happens and come with many biases.

Furthermore, our cognitive structures themselves may well be altered by this information overload in this data explosion age: more than a temporary attraction of our attention, we run the risk of impairing our ability over the long term to remember, focus and process information, with children being particularly vulnerable to these changes.[32] We could be the victims of a civilisation in which our mental health is in danger after exposure to social media and the internet for more than 30 consecutive minutes, a civilisation in which our attention span on a given subject is reduced to a few seconds, making each of us a goldfish trapped in the fishbowl of our screens. This is the picture Bruno Patino paints in his influential book, which leads to the question Nicholas Carr raises: "Is Google Making Us Stupid?"[33]

Furthermore, several neurological studies have revealed the real limits of our multitasking abilities, highlighting in addition how our attention is disrupted ever more easily by external distractions.[34]

Technological solutions and recommendations

At a time when we are experiencing what we might call "digital Diogenes syndrome", with every person and every organisation storing as much data as possible, remedies to this "syllogomania 2.0" are essential The hunt for technological solutions to prevent and contain the data explosion and information overload must be stepped up.

This includes research into data aggregation and curation in particular. This involves selecting, editing and sharing only relevant content:[35] using more effective ways to select relevant data could limit the amount of data stored. There are also several database projects investigating how to erase some data gradually using "controlled decay" technology.[36] Another example is Personal Information Management Systems (PIMS), a technology that is still in its infancy but which could provide users with the means to manage and control all their personal data. Broadly speaking, Big Data-based computing holds enormous potential. Over the next few years, major breakthroughs are expected that will improve the data management and data processing aspects of digital technologies: advances in computing and storage architectures, data integration and curation, models, software, algorithms, artificial intelligence, acquisition and visualisation systems, and quantum technologies.[37]

But digital sufficiency is first and foremost a cultural struggle. This is why training in digital issues is essential. The goal is not so much to learn how to use these tools as to understand their consequences and the legislation that governs them. For example, despite the General Data Protection Regulation (GDPR), which governs the processing of personal data in the European Union, only 6% of French people refuse cookies on websites.[38] Data protection legislation could be strengthened or at least backed up by improved education. There is a need to promote digital sufficiency and develop digital hygiene at both the individual and organisational levels (companies and public authorities), which would include regular data clean up days.

Finally, the desire for digital sovereignty and independence from large digital companies requires an open debate on the monopoly wielded by these – mostly American – platforms, and on the algorithms that capture our attention or create filter bubbles that trap us. This goal questions whether French and European governance of data and its infrastructure is feasible. Several of the people interviewed argued that the last resort was to simply disconnect,[39] a response that remains stubbornly uncommon in our society.

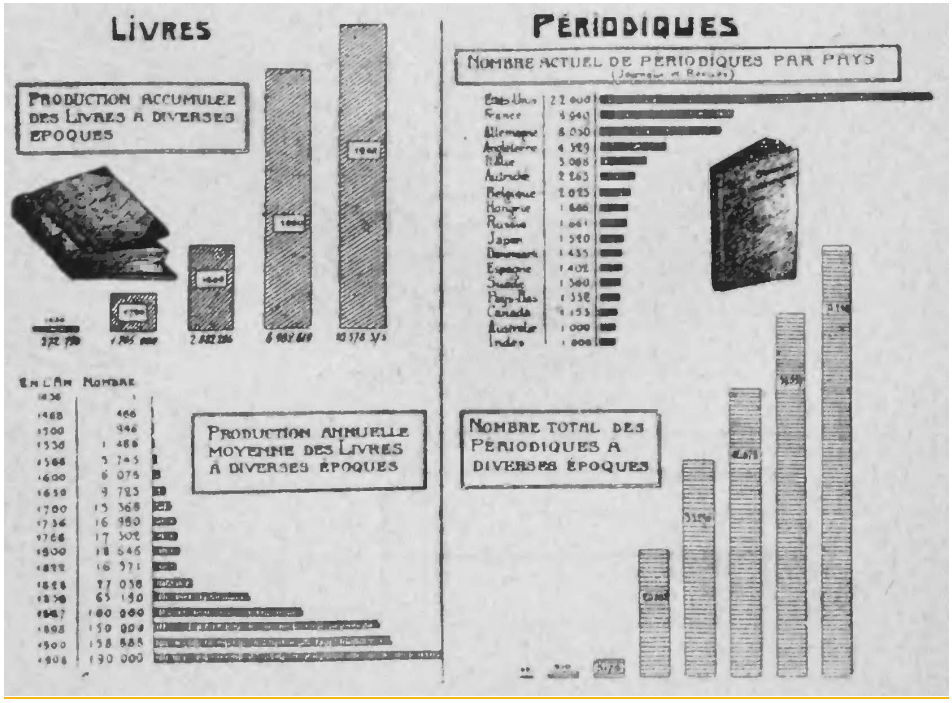
The Office's websites:
http://www.assemblee-nationale.fr/commissions/opecst-index.asp
http://www.senat.fr/opecst

Références

[1] *According to Robert Darnton, "Every age was an age of information, each in its own way". See "An early information society: News and the media in eighteenth-century Paris", American Historical Review, no. 1, vol. 105, 2000. See https://academic.oup.com/ahr/article/105/1/1/64445*

[2] *Barnaby Rich, writing in 1613, complained of the proliferation of literature:* "One of the diseases of this age is the multiplicity of books; they doth so overcharge the world that it is not able to digest the abundance of idle matter that is everyday hatched and brought forth into the world." *Derek de Solla Price once referred to a "Barnaby Rich effect" to describe information overload. Similarly, in 1621, Robert Burton said:* "What a glut of books! Who can read them? As already, we shall have a vast chaos and confusion of books; we are oppressed with them, our eyes ache with reading, our fingers with turning." *In 1934, in his* Treatise on Documentation*, Paul Otlet highlighted the issue of documentary inflation, observing with regard to books and documents that* "their enormous mass, accumulated in the past, increases every day, every hour, by disconcerting, sometimes alarming numbers. […] From them like water falling from the sky, we can say that they can cause flooding and flood or spread to become beneficial irrigation" *(see also his* Manuel de la bibliothèque publique*, 1922.)*

How the data explosion was perceived a century ago



Source: Paul Otlet, Public Library Manual*, 1922.*

More recently, Emmanuel Didier observed how frequently people talk about data explosion: "There is a very beautiful thing that strikes you as a historian of statistics, and that is that, since at least 1850, every time there is a technological revolution, there is talk of us being overwhelmed by data" *(see his comments recorded by Edouard Mien and Alizé Papp, "L'afflux massif de données est-il nouveau ?", Regards croisés sur l'économie, 2018, no. 23). The reason why the previous data explosions no longer cause us concern and the glut of books is no longer perceived as a data deluge is mainly because information management tools have provided a means of channelling this information overload. Pascal Griset and Bruno Strasser explained this during their hearings, citing examples such as encyclopaedias and the Memex (memory extender), Vannevar Bush's 1945 concept that was the precursor to computerised systems for managing databases.*

[3] *Statista's data is based on a report by IDC, which, as well as forecasts on the amount of digital data created and copied, provides others on the distribution of companies using data, the possible benefits of this data and data storage, for example. It is important to note that only a small proportion of the digital data created is stored: only 2% of the data produced in 2020 was still stored in 2021. See their regularly updated website https://www.statista.com/ and, in particular, this graph on data growth https://www.statista.com/statistics/871513/worldwide-data-created/ https://fr.statista.com and the report by David Reinsel, John Gantz and John Rydning,* The Digitization of the World*, 2018, link: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf*

[4] *OPECST has already addressed the issue of data storage, its current challenges and the solution of storing data in the form of DNA, in Ludovic Hayes Science and Technology Briefing "Storing data in the Form of DNA", no. 29, December 2022, available at the following links: on the Senate website www.senat.fr/rap/r21-285/r21-2851.pdf or on the National Assembly website https://www2.assemblee-nationale.fr/content/download/451536/4391889/version/1/file/OPECST_2021_0063_Note_stockage_ADN.pdf*

[5] *Quotation from Daniel Rosenberg's book,* Raw Data is an Oxymoron*, 2013.*

[6] *Data can first be categorised according to its* ***content****, either quantitative data (a number) or qualitative data. Qualitative data is subsequently separated into two further categories: nominal data (which cannot be sorted, such as eye colour) or ordinal data (which can be sorted, such as clothing size). It is also possible to categorise data according to its* ***format****, with structured data (subject to a fixed format depending on a model) being distinguished from unstructured data (e.g. images, audio or video whose form is not known in advance), although there is also semi-structured data, which falls somewhere between structured and unstructured data (it has no formal structure but still has organisational properties, often in the form of metadata, which consists of* "descriptions attached to data or documents, intended to make it easier to categorise them and provide access to their information content. They can be used to name, describe, catalogue and denote intellectual property or copyright", *see Mokrane Bouzeghoub and Rémy Mosseri (eds.),* Les Big Data à découvert, *CNRS éditions, 2017). A distinction is made between data lakes, which can hold any type of data, especially unstructured data, and data warehouses, which hold structured, transformed and cleaned data. It is also possible to categorise data according to its* ***frequency of use****, distinguishing between hot data, which is used very frequently (and which therefore requires storage media that can be accessed quickly or immediately), and cold data, which is accessed only rarely, if at all. Lastly, data can be categorised according to its* ***source****:* first-party *data is data directly retrieved from the company's users or customers,* second-party *data is data purchased from the company that collected it, and* third-party *data is data purchased from a company that did not collect it itself. Data categorisation processes are crucial for more effective data management and protection.*

[7] *In 2017, the OPECST explored artificial intelligence technologies and the issues surrounding them in depth in Report No. 464 (2016-2017) by Mr Claude de Ganay, Member of the National Assembly, and Ms Dominique Gillot, Senator, "Toward a Controlled, Useful and Demystified Artificial Intelligence". See the links available on the Senate website http://www.senat.fr/notice-rapport/2016/r16-464-1-notice.html and on the National Assembly website https://www.assemblee-nationale.fr/dyn/14/dossiers/intelligence_artificielle_maitrisee_utile*

[8] *In his book on Big Data, Pierre Delort gives the example of Google Flu Trends, which, together with the Centers for Disease Control and Prevention (CDC), examined the correlation between Google searches containing flu-related terms and the number of people infected with flu between 2008 and 2015. Its convincing results were announced two weeks before the epidemic was detected. This example of an opportunity presented by data in the field of health has its limitations, however, since it also measures people's reactions to media coverage of flu or to rumours: when people search for information about a disease, it does not mean they are sick. See Pierre Delort,* Le Big Data*, PUF, 2018. More recently, the COVID pandemic prompted discussion on the relevance of analysis factors and the quality of indicators. These examples suggest that caution be exercised when interpreting data. More generally, the use of data in health must be guided by scientific and ethical principles, as the Spanish Parliamentary Office for Science and Technology points out in a short report, published in November 2022, that looked at the relationship between artificial intelligence and health. See https://www.oficinac.es/sites/default/files/informes/20221114_Report%20C%20IA_0.pdf*

[9] *A paper from the UK Parliamentary Office of Science and Technology (POST) explains how data is used in transport today and the opportunities it offers, in particular the potential for optimising transport and the future widespread use of connected cars. It also examines future challenges such as public access to data, data aggregation between all stakeholders in the transport sector and data governance. See the August 2014 POST briefing, "Big and Open Data in Transport": https://post.parliament.uk/research-briefings/post-pn-472/*

[10] *As Serge Abiteboul and Valérie Peugeot explain, these sites often offer "free" services that are not truly free:* "users provide in exchange, not money, but their attention and 'info-currency', information about themselves". *This is coupled with a* "monopolistic or oligopolistic phenomenon on a global scale": *a small number of large companies are monopolising the lion's share of personal data, which gives them a commercial advantage, hampers innovation, accentuates the increasingly unequal relationship between users and these companies, weakens personal data protection and raises the issue of how to harness this informational infrastructure for surveillance purposes. See Serge Abiteboul and Valérie Peugeot,* Terra Data. Qu'allons-nous faire des données numériques ?, *Le Pommier, 2017.*

[11] *Several Big Data-related careers have emerged with the arrival of this data explosion. These jobs revolve around the different stages involved in working with Big Data: data collection, data cleansing, data analysis and the presentation of results. Here are some examples:* data engineer, data steward, data architect, data scientist, data analyst, data

visualisation and data protection officer. *The demand for these skills is growing, and training courses are already beginning to appear. Serge Abiteboul suggests ways of training future data scientists and creating new training courses (see Serge Abiteboul et al., "L'émergence d'une nouvelle filière de formation : 'data scientist'", INRIA Saclay, CNAM-Paris, CEDRIC laboratory, 2014).*

[12] *The figures for GHG emissions from digital technology are based on the work of Hugues Ferreboeuf (report "Towards Digital Sufficiency", The Shift Project, 2018, updated in 2021: https://theshiftproject.org/wp-content/uploads/2019/03/Lean-ICT-Report_The-Shift-Project_2019.pdf); the assessment of global primary energy consumption was carried out by Frédéric Bordage (see "Empreinte environnementale du numérique mondial", GreenIT.fr, 2019: www.greenit.fr/empreinte-environnementale-du-numerique-mondial/?msclkid=6b823725ba7211ec92840260f951396f). This data is included in the joint report by ADEME and ARCEP, "Évaluation de l'impact environnemental du numérique en France et analyse prospective", 2022: https://www.arcep.fr/uploads/tx_gspublication/etude-numerique-environnement-ademe-arcep-note-synthese_janv2022.pdf . The data life cycle has* ***four stages*** *that consume natural resources and emit greenhouse gases (GHGs). First, metals are mined to manufacture devices such as connected objects, sensors, smartphones and servers. All these devices then have to be manufactured. The data is then transmitted, processed, stored and combined with other data. Lastly, resources are consumed during the use phase of these devices. (See the EcoInfo group's book, coordinated by Françoise Berthoud,* Les impacts écologiques des Technologies de l'Information et de la Communication*, EDP Sciences, 2012.) Manufacturing hardware, computer centres and networks is responsible for 44% of the carbon footprint of digital technology worldwide, and using them for 56%. A Citizing study estimated that, in France, digital technology would account for 2% of total GHG emissions in 2019. The study shows that computing devices (smartphones, laptops, etc.) are responsible for a very high proportion of the environmental impacts caused by digital technology (81%). Manufacturing and distributing these devices alone generates 86% of their total emissions and is therefore responsible for 70% of the total carbon footprint of digital technology in France. This is much higher than the worldwide figure of 44% due to the import of digital equipment from Asian countries, where the carbon intensity of electricity is much higher than in France (see the Senate Information Report by Guillaume Chevrollier and Jean-Michel Houllegatte on behalf of the Senate Regional Planning and Sustainable Development Committee, "Pour une transition numérique écologique", No. 555, 2019-2020: www.senat.fr/rap/r19-555/r19-555.html). The low level of recycling of connected objects is another factor that contributes to these environmental impacts as it is likely to be a major new source of electronic waste (74 million tonnes per year by 2030), with 5.3 billion mobile phones being thrown away in 2022 alone (see the 2022 reports of the International Forum on Waste Electrical and Electronic Equipment and the Global E-waste Monitor, along with the earlier report by Thomas Graedel (ed.), "Recycling rates of metals, a status report", UNEP, International Resource Panel, 2011: www.unep.org/resources/report/recycling-rates-metals-status-report?msclkid=eff8f215ba7211ec88247081146be314). Furthermore, as Serge Abiteboul pointed out during his hearing, there is a tendency to focus on the environmental impact of data centres alone, whereas the sheer volume of digital data in circulation should not be overlooked. Videos are an important part of the problem, especially high- or very-high-definition videos.*

[13] *The expression "new capitalist model" is used in an article by Alizé Papp "L'infobésité, une épidémie à l'âge des nouvelles technologies de l'information et de la communication" (Regards croisés sur l'économie, no. 23, 2018, see https://www.cairn.info/article.php?ID_ARTICLE=RCE_023_0105). And Yann Moulier-Boutang has primarily theorised the concept of cognitive capitalism, the third stage of capitalism after its origins in mercantilism and its industrial form (see* Le capitalisme cognitif : la nouvelle grande transformation*, Éditions Amsterdam, 2007). Other authors have written about the attention economy, including Yves Citton in the book he edited (see* L'économie de l'attention, nouvel horizon du capitalisme ?*, La Découverte, 2014.). In 1969, Herbert Simon pointed out:* "In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it" (see *Herbert Alexander Simon,* The Science of the Artificial*, MIT Press, 1969, translation by Jean-Louis Le Moigne).*

[14] *See Eli Pariser, who coined the phrase,* The Filter Bubble: What the Internet Is Hiding from You*, Penguin Press, 2011.*

[15] *See Hervé Le Crosnier, "De l'information à l'influence", Documentation et Bibliothèques journal, special issue "Quelle société de l'information ? Pour quelles bibliothèques/services d'archives ?", volume 64, no. 4, 2018.*

[16] *Julia Cagé, Nicolas Hervé and Marie-Luce Viaud show that 64% of the information published online is simply copied and pasted. In their study, they not only looked at the quantity of information copied, but also at the speed at which this information is passed from one site to another. For straightforward news repeats,* "half of the events propagate in less than 25 minutes; a quarter in just […] 230 seconds, and 10% in just four seconds" *(see Julia Cagé, Nicolas Hervé and*

Marie-Luce Viaud "L'information à tout prix", Questions de communication, no. 32, 2017). The problem is not entirely new, since the Havas Agency introduced a system for using carrier pigeons and was at the forefront of the use of the telegraph. However, according to Alizé Papp, this frenetic pace is accelerated by digital civilisation, which results in the standardisation of content and a decline in the quality of what is offered. Furthermore, this growing problem "would lead to the spread of infobesity by making people accustomed to reading homogenised and poorly digested content, i.e. a form of informational junk food" *(see her article "L'infobésité, une épidémie à l'âge des nouvelles technologies de l'information et de la communication", Regards croisés sur l'économie, no. 23, 2018).*

[17] *On this subject, Amaël Cattaruzza develops in his book the idea that the digital space is a "geo-political" space in which we need to be concerned about the domination of American firms. Although some countries are trying to build their own data centres to re-establish their digital sovereignty, i.e. to keep data on their territories, France suffered the failure of the Andromeda project in 2011 (see* Amaël Cattaruzza, Géopolitique des données numériques, *Pouvoir et conflits à l'heure du Big Data, Le cavalier Bleu, 2019). On the concept of "surveillance capitalism", see the following references: Mark Andrejevic, "Surveillance in the Digital Enclosure", The Communication Review, vol. 10, no. 4, 2007; John Bellamy Foster and Robert McChesney, "Surveillance Capitalism. Monopoly-Finance Capital, the Military-Industrial Complex, and the Digital Age", Monthly Review, vol. 66, no. 3, 2014; Viktor Mayer-Schönberger and Thomas Ramge,* Reinventing Capitalism in the Age of Big Data*, Basic Books, 2018; Virginia Eubanks, Automating Inequality: How High-tech Tools Profile, Police and Punish the Poor, St. Martin's Press, 2018; Shoshana Zuboff,* L'âge du capitalisme de surveillance*, Zulma, 2020; Cédric Durand,* Techno-féodalisme. Critique de l'économie numérique*, La Découverte, 2020; Charleyne Biondi,* Dé-coder. Une contre-histoire du numérique*, Bouquins, 2022.*

[18] *See for example https://www.lemonde.fr/pixels/article/2022/12/23/l-entreprise-chinoise-bytedance-proprietaire-de-tiktok-admet-avoir-espionne-des-journalistes_6155484_4408996.html*

[19] *See Eszter Hargittai, "Second-Level Digital Divide? Differences in People's Online Skills?", First Monday, no. 7 (4), 2002; Eszter Hargittai and Paul Di Maggio, "The New Digital Inequality? Social Stratification among Internet Users?", presentation at the annual conference of the American Sociological Association, Chicago, 2002; the report of the French National Digital Council, "Citoyens d'une société numérique : pour une nouvelle politique d'inclusion", 2013; or Fabien Granjon et al.,* Inégalités numériques. Clivages sociaux et modes d'appropriation des TIC, *Hermès/Lavoisier, 2008. According to Francis Jauréguiberry, there are the "new telecommunications-poor" whose lives are governed by an obligation to respond immediately to emergencies and who therefore cannot disconnect and escape situations where they have to respond continually to the demands made on them.*

[20] *See Caroline Sauvajol-Rialland,* Infobésité, Comprendre et maîtriser la déferlante d'informations*, Vuibert, 2013.*

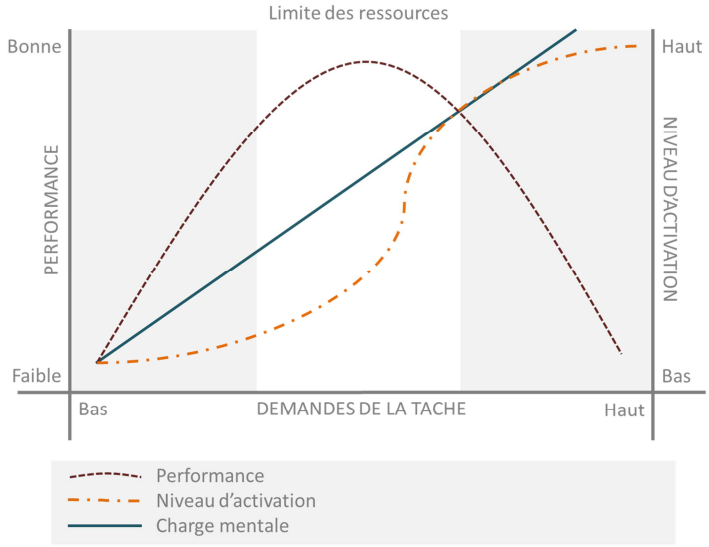
[21] *Quotation from* The Rock*, T.S. Eliot, 1934. Not all specialists agree on separating the concepts of information and data in this way, but many authors and companies do make this distinction. For some of them, it is the very storage of data that makes no sense, firstly because too much data makes it difficult to identify which data is relevant, and secondly because data has no meaning in itself, unlike information (see Thomas H. Davenport and Laurence Prusak,* Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, 1998).*

[22] *See for example Éric Campoy, Michel Kalika and Henri Isaac "Surcharge informationnelle, urgence et TIC. L'effet temporel des technologies de l'information", Management & Avenir, vol.3, no. 13, 2007.*

[23] *See the joint report by the French Centre for Strategic Analysis and Directorate General for Labour, coordinated by Tristan Klein, "L'impact des TIC sur les conditions de travail", 2012, available at the following link: http://archives.strategie.gouv.fr/cas/system/files/raptic_web_light_final28022012.pdf*

[24] *Working memory, a type of current, short-term memory that allows us to store and manipulate information for a short period of time for use in completing a task, can only handle a limited number of pieces of information at a time, about seven, plus or minus two (see one of the most cited articles in psychology, George Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information", Psychological Review, vol. 63, no. 2, 1956). Research over the last decade or so has reduced this to three or four pieces of information (see Jeanne Farrington, "From the Research: Myths Worth Dispelling. Seven Plus or Minus Two", Performance Improvement Quarterly, vol. 23, no. 4, 2011). Among the studies that have examined the performance of individuals in relation to the mental workload involved is the meta-analysis by Mark S. Young (See Mark S. Young et al., "State of Science: Mental Workload in Ergonomics", Ergonomics, vol. 58, no. 1, 2015), on which Rémi Mounier based the chart below. One observation is that around a maximum level of performance for a given mental workload, productivity decreases rapidly above and below it. However, the characteristics of this curve obviously also depend on the individual and the task assigned.*

Optimal performance according to mental load



Source: Rémi Mounier's Octopus Ergonomie blog, article on "La surcharge cognitive (ou mentale)",https://octopus-ergonomie.com/blog-surcharge-mentale-cognitive-15

[25] *Phrase used by Michel Kalika during his hearing.*

[26] *For some authors, such as David Shenk, Caroline Sauvajol-Rialland and Michel Kalika, the issue is not only linked to information overload, but also to communication overload, which in turn reinforces the feeling of information overload (see David Shenk,* Data Smog: Surviving the Information Glut*, Harper & Collins, 1997, and Caroline Sauvajol-Rialland,* Infobésité, comprendre et maitriser la déferlante d'informations*, Vuibert, 2013). Although 90% of employees say they receive too many unnecessary emails and 74% of managers say they suffer from information overload as a result of a widely acknowledged sense of urgency and time pressure, this is a problem that affects organisations as well as individuals. According to IBM, large companies are unable to interpret up to 90% of the data they generate. This uninterpreted data is referred to as* Dark Data. *However, some have criticised the idea of a link between Big Data and infobesity, as Marie-Anne Chabin has said:* "infobesity can exist without Big Data; and Big Data does not necessarily cause infobesity". *According to her, Big Data is an objective reality, born of technological progress, whereas infobesity is a subjective concept, where the problem lies with the user and not with the existence of the data itself. See her article "Infobésité et Big Data : ne pas confondre, Documentaliste – Sciences de l'information journal, Vol 51, 2014.*

[27] *According to Caroline Sauvajol-Rialland, ICT has brought about several changes that increase information overload: firstly, work that is increasingly abstract and interactive, work that is transparent and monitored by management, work whose pace is accelerating, work that is increasingly carried out over email without this reducing the amount of face-to-face communication required, and, sometimes, remote work that brings with it space management issues. Lastly, on a broader level, ICT means that work is ever changing and new skills are in constant demand.*

[28] *However, Caroline Sauvajol-Rialland emphasises that emails also make work easier and that the speed of decision-making is generally faster as a result of email systems:* "the burdensome effects of email are more the result of mistaken and/or abusive use of the tool's functionalities than of the tool itself" *(see her book, op. cit.). There are several factors that may explain this misuse of email, including the time required to respond and the risk of redundancy in an email thread's running list of all succeeding replies.*

[29] *See Martin Kersten and Lefteris Sidirourgos "A Database System with Amnesia", CIDR, 2017: https://www.cidrdb.org/cidr2017/papers/p58-kersten-cidr17.pdf*

[30] *According to Chris Anderson, the challenge that Big Data poses to traditional scientific methods could lead to the end of theory, in other words, to the end of science as we know it. See Chris Andersonn, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", June 2008: https://www.wired.com/2008/06/pb-theory/*

[31] *Alain Desrosières use of this phrase shows that, in general, data is not automatically accessible; it has to be retrieved. This in turn implies that a choice is made about the data retrieved, the statistical models used, etc. Biases may therefore be introduced during these construction operations. See Alain Desrosières,* La politique des grands nombres, Histoire de la raison statistique*, La Découverte, 2010.*

[32] *In addition to children's overexposure to screens, which makes them dependent on digital tools from an early age, and the effects – specific to them – of the information overload they experience, there are cognitive and physical*

impacts, especially when younger children are deprived of interaction with their parents, who are themselves overly absorbed by their screens. These issues were raised about ten years ago by psychiatrist Serge Tisseron, philosopher Bernard Stiegler and anthropologist Pascal Plantard. Their implications prompted the French Academy of Sciences and subsequently the French Academy of Medicine and Academy of Technologies to take action on the matter, resulting in some particularly worrying in-depth empirical studies, such as Marie Claude Bossière's work. See in particular the following: Serge Tisseron and Bernard Stiegler, Faut-il interdire les écrans aux enfants ?, *Mordicus, 2009; Pascal Plantard, "Numérique et éducation : encore un coup de tablette magique ?", Administration et Éducation, no. 146, 2015, "Le collège et les pratiques numériques des adolescents", Les Cahiers Pédagogiques, no. 520, 2015, with Jonathan Bernard and Sophie Jehel "Tablette, smartphone, console, télé, ordi... Faut-il les interdire aux enfants ?", INSERM journal, 2019; Jean-François Bach, Olivier Houdé, Pierre Léna and Serge Tisseron,* L'enfant et les écrans, *opinion of the French Academy of Sciences, Le Pommier, 2013; "L'enfant, l'adolescent, la famille et les écrans", joint appeal of the French Academy of Sciences, Academy of Medicine and Academy of Technology, 2019, https://www.academie-sciences.fr/pdf/rapport/appel_090419.pdf; Marie-Claude Bossière, "Le pédopsychiatre et la toxicité de l'omniprésence des écrans", Nouvelle Revue de l'Enfance et de l'Adolescence, vol. 2, no. 1, 2020, with Daniel Marcelli and Anne-Lise Ducanda, "L'exposition précoce et excessive aux écrans (EPEE) : un nouveau syndrome", Devenir journal, vol. 32, no. 2, 2020.*

[33] *See Bruno Patino,* La civilisation du poisson rouge : Petit traité sur le marché de l'attention*, Grasset, 2019, and Nicholas Carr, "Is Google Making Us Stupid? What the Internet is Doing to Our Brains", The Atlantic, no. 7, 2008, available at: https://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/. In his book, Bruno Patino explains how Google engineers have managed to calculate the maximum attention span of a goldfish that swims in circles around in its bowl and seems to rediscover the world with each lap (eight seconds). He then points out that these same engineers have assessed the attention span of so-called millenials, who reached adulthood around the year 2000 and grew up using internet-connected screens (nine seconds). This is why, according to him, we have become* "goldfish, trapped in the fishbowl of our screens, subjected to the whirlwind of notifications and instant messages, totally dependent on technologies that are dangerous because they can lead to solitude and depression". *Patino cites well-known study that estimates that 30 minutes is the maximum time that people should be exposed to social networks and internet-connected screens. Any longer is a threat to our mental health. See Melissa G. Hunt, Rachel Marx, Courtney Lipson and Jordyn Young, "No More FOMO: Limiting Social Media Decreases Loneliness and Depression", Journal of Social and Clinical Psychology, December vol. 37, no. 10, 2018.*

[34] *Diego Mac-Auliffe Cabello writes in his doctoral thesis:* "the analysis of behavioural performance clearly shows that performing two tasks simultaneously reduces overall performance. Our iEEG results indicate that both tasks require the same cognitive resources – that is, simultaneous activity of the same neural populations in the above-mentioned regions – that cannot be allocated at the same time. The executive system therefore forces the two tasks to be performed one after the other rather than at the same time, resulting in an immediate decrease in response speed and an increase in errors" *(see Neural mechanisms underlying external distraction by unexpected environmental stimuli or by a secondary task: an intracranial EEG investigation, University of Lyon, 2020). Similarly, Jean-Philippe Lachaux, in his study of the brain's reaction to distractions, concludes with the idea that there are several types of "attention": voluntary attention (when we choose to pay attention), reflexive attention (for example, when an oncoming car could run you over) and attention based on emotion, linked to our reward system (for example, when we receive a text message or notification of a "like" on Facebook). It is a commonly held belief that we tend to exaggerate the importance of voluntary attention; however, if we are in fact always or almost always paying attention, the question is to what.*

[35] *Marc Rougier, founder and president of the curation platform Scoop.it France, says that curation is* "the convergence of three components that involve selecting, organising or editing, and then sharing existing content". *These new technologies* "provide a means of sorting through the excess of information and focusing on what is relevant", *according to Steven Rosenbaum, thereby providing a response to infobesity. However, this technology also has its detractors, who accuse it of* "encouraging the plundering of the web (by the web) through the abuse of copy-and-paste" *(see Caroline Sauvajol-Rialland,* Infobésité, comprendre et maitriser la déferlante d'informations*, Vuibert, 2013). Nevertheless, growth in these content curation technologies is booming, with many tools and platforms already available or under development, both for individuals, such as Pearltrees, and for companies, such as Knowledge Plaza.*

[36] *See Martin Kersten and Lefteris Sidirourgos "A Database System with Amnesia", CIDR, 2017: https://www.cidrdb.org/cidr2017/papers/p58-kersten-cidr17.pdf*

[37] *According to Rolf Landauer's famous maxim, information is physical in nature. A quantum information medium must therefore allow the transmission and quantum processing of this information, using a different set of rules from those understood in the past and with new calculation or cryptographic methods that could be more effective in managing immense quantities of data. This perspective is attracting interest from many mathematicians, physicists, computer*

scientists and biologists, despite the technological obstacles and barriers to understanding. In 2019, Cédric Villani, then President of OPECST, presented a series of science and technology briefings on these technologies (briefings No. 13 "Quantum Technologies: Introduction and Issues", No. 15 "Quantum Technologies: Quantum Computing", No. 16 "Quantum Technologies: Quantum Programming" and No. 18 "Quantum technologies: Quantum and Post-Quantum Cryptography") available at the following links: http://www.senat.fr/opecst/notes.html or https://www2.assemblee-nationale.fr/15/les-delegations-comite-et-office-parlementaire/office-parlementaire-d-evaluation-des-choix-scientifiques-et-technologiques/(block)/48190

[38] *According to a study published in 2022 by NordVPN, only 6% of French people refuse all cookies, with the vast majority of users accepting them just to save time, even though most users know that they should be wary of them (see https://www.clubic.com/pro/legislation-loi-internet/donnees-personnelles/actualite-424824-vie-privee-seuls-6-des-francais-refusent-tous-les-cookies.html and https://siecledigital.fr/2022/05/31/46-pc-internautes-france-acceptent-tous-les-cookies/). A slightly less recent IFOP survey for CNIL found that around 22% of French people refuse cookies when websites ask for their consent, see Romain Bendavid and Delphine Poët, "Les Français et la règlementation en matière de cookies", IFOP study for CNIL, 2019, available at https://www.cnil.fr/sites/default/files/atoms/files/les_francais_et_la_reglementation_en_matiere_de_cookies_-_sondage_ifop_pour_la_cnil_-_decembre_2019_.pdf*

[39] *Several specialists in digital issues emphasised the idea of resorting to disconnection during their hearings, and Ioana Manolescu even remarked that* "Steve Jobs' children were brought up without computers". *Article L 2242-17 of the French Labour Code, resulting from French Law 2016-1088 of 8 August 2016 on work, the modernisation of social dialogue and the securing of professional careers, enshrined the right to disconnect in law. The question of potential data cleansing obligations remains open, as does the subject of those activities where reducing or preventing the production of data would be relevant.*

Persons consulted

Institutions

French Academy of Sciences
- Mr Serge Abiteboul, researcher at the National Institute for Research in Digital Science and Technology (INRIA), former professor of the Computer Science Chair at the Collège de France, member of the French Academy of Sciences and of the College of the French Regulatory Authority for Electronic Communications, Postal and Print Media Distribution (ARCEP), former member of the French National Digital Council, former chairman of the Scientific Council of the Computer Society of France

French Academy of Technology
- Mr Gérard Roucairol, honorary chairman of the French Academy of Technology, chairman of the Academy's digital unit, former scientific director of the Bull Group, former professor at the École Normale Supérieure and at the University of Orsay
- Mr Claude Le Pape-Gardeux, member of the French Academy of Technology, member of the Scientific Council of the Institut Mines-Télécom, former member of the Scientific Council of CNRS, former company executive

French Government
- Ms Laure Lucchesi, director of ETALAB, Interministerial Directorate of Digital Affairs, Prime Minister's Office
- Ioana Manolescu, research director at the National Institute for Research in Digital Science and Technology (INRIA) and at the École Polytechnique, scientific director of the AI Lab, ETALAB
- Mr Paul-Antoine Chevalier, head of the data exploitation unit, AI Lab, ETALAB

Researchers

French National Centre for Scientific Research (CNRS)
- Mr Pierre-Éric Mounier-Kuhn, researcher at CNRS, historian of computer science

Universities
- Ms Anastasia Ailamaki, computer scientist, engineer and full professor at the Swiss Federal Institute of Technology Lausanne (EPFL), co-founder of the start-up Raw Labs, specialising in data management software
- Mr Mokrane Bouzeghoub, Professor Emeritus at the University of Versailles, coordinator of scientific actions on data management within the CNRS Mission for Interdisciplinarity (MITI)
- Pascal Griset, Professor of Contemporary History at the University of Paris-Sorbonne, historian of innovation, president of the committee for the history of the National Institute for Health and Medical Research (INSERM)
- Mr Francis Jaureguiberry, Professor of Sociology at the University of Pau, member of the Passages laboratory, coordinator of the "DEconnexion Volontaire des Technologies de l'Information et de la Communication" (DEVOTIC) project
- Mr Michel Kalika, Professor Emeritus at the IAE Lyon School Of Management, Chairman of the Business Science Institute and of the University of Jean-Moulin-Lyon-II, former Director at the École de Management (EM) Strasbourg, former Professor at the University of Paris Dauphine
- Ms Caroline Sauvajol-Rialland, lecturer at the Paris Institute of Political Studies (Sciences Po), at the International Institute for Business and Development (ICD) and at the Catholic University of Louvain
- Ms Valérie Schafer, Professor of Contemporary European History at the University of Luxembourg, historian of information technology
- Mr Bruno Strasser, Professor at the University of Geneva and Yale, Director of The Bioscope, a public laboratory of the life and biomedical sciences at the University of Geneva

Companies

Orange
- Valérie Peugeot, head of Future Studies at Orange, lecturer at the Paris Institute of Political Studies (Sciences Po), member of the College of the French National Commission for Data Protection (CNIL) and former vice-president of the French National Digital Council (CNNum)

Shopify
- Ms Alizé Papp, data scientist

Dataiku
- Mr Léo Deyfus-Schmidt, Vice President for Research
- Mr Rémi Meunier, Vice President – Public Sector
- Mr Vivien Tran Thien, Director of Artificial Intelligence

Associations

The Shift Project
- Mr Hugues Ferreboeuf, Director of Virtus Management, polytechnician, engineer of the Corps des Mines, responsible for digital sufficiency issues at The Shift Project