

Call for multidisciplinary contributions on:

“Big data and social protection”

Issue 4 - 2017

This call for contributions is meant for researchers in the fields of applied mathematics, computer science, health economics, economics and econometrics, law, philosophy and sociology as well as those who work within the social protection field.

Articles are to be submitted before **March 31st, 2017**.

Contemporary society is undergoing a rapid transformation, one of the distinctive aspects of which is the digitalization of most sectors of the economy. Firms and network activities (social networks, digital platforms, and connected objects) generate and store ever increasing amounts of data, and require more automated, complex and faster processing. Data is therefore at the heart of value creation in the digital field and, more generally, in the economy. New stakeholders (Google, Facebook, Uber, Airbnb just to name a few) emerge and aggressively shake long-standing economic models. Certain sectors which interact with the social field are experiencing, or will experience in the near future, particularly profound changes, especially healthcare, housing, mobility, insurance, finance or education. The generalized exploitation of this growing mass of data creates new knowledge, services and social organizations. It is often called “Big data”, which denotes the processing of large amounts of old and new types of data.

The growing quantity and variety of stored data along with increasingly powerful digital processing capabilities have therefore lead to the creation of a new domain of research and applications, at the border between statistics and computer science. Big data covers certain characteristics of data as much as it does to the methods used to exploit it and the end goals at stake. Its definition and boundaries are amorphous and ever-changing but they are also sufficiently present in public discourse (colloquiums, written works, reports...) and in

academic works (beyond the numerous articles published in thematic reviews, Big data even has an English-language¹ review in order to ensure that our review would assign it a file. In our daily lives in France and elsewhere, Big data also intervenes through generating personalized advertisements by sweeping inboxes, localizing Uber cars, surveillance of public places, detecting the development of diabetes or advance orientation in urgent services thanks to the use of hospital data (Israel), the detection of allocation fraud, obtaining building credits, etc.

The concept of Big data emerged at the end of the 1990s and greatly expanded at the end of the 2000s, when it came to embody its most expansive definition, that of the “3 Vs”: volume, variety and velocity. We will discuss it here in its most widely accepted form, certainly relying on the intersection of different kinds of data with different original purposes (“variety”), which are often (not systematically) voluminous but it will also include a range of analytical and processing techniques that come from statistical learning.

We will be focusing on two broad categories of Big data use as they emerge through current projects: one category includes data use geared towards research and information gathering which explains behavioral patterns, inequalities, health status, etc; and another category, which is rapidly expanding now, related to individual cases, to facilitate decision-making: from job hunting, diagnosis and strategies for therapy.

Big data is polymorphous; it includes “structured” traditional data in the field of human and social sciences, like administrative files (unemployment insurance data, hospital and illness insurance data, data for managing allocations for family assistance, etc.) or investigative data. Their newness comes from the fact that they also include data that is considered “non-structured” like textual data, imaging data, metadata, social network data, forum data, etc. The field of data on which health and social protection research can rely is therefore experiencing major growth, and is enriched by data produced outside of the social protection system: we are referring here to data produced by new private actors (most notably through connected objects in healthcare), but also to data produced in completely different contexts and for reasons that do not necessarily tie in with health or social security (emails sent by consumers or users, supermarket purchases, passages close to mobile phone antennae, etc.). Certain data that is useful to Big data is considered open data (that which is or could be freely accessible because it is not protected by law: private life secrets, business secrets...); others are protected by laws, which can only be circumvented under certain conditions and for certain reasons. This could be a particularly interesting topic for legal articles.

This call for papers casts a wide net over the issues linked to Big data in the field of the RFAS: new statistical techniques, new individual and collective uses, new regulations. The legal and ethical issues raised will be given special attention. Interesting perspectives gleaned through foreign experiences will be welcomed as well.

Theme 1 : Classical statistics and statistical learning

Big data covers a variety of statistical methods that have long been known but which have only recently come back into focus due to the existence of large amounts of data and technical capabilities that have become available to process them. Researching weak signals, detecting

¹ <https://journalofbigdata.springeropen.com/>

profiles, neuron networks...would come to supplement (or even replace?) classical statistics. Big data is notably characterized by its ability to update unknown correlations and essentially seeks to observe them and predict events (which can pose problems when these correlations are not steady over time) and not to explain them. So then what are the specificities of Big data compared to “traditional” methods? Do gateways not exist between classical statistics, which relies on the definition of a model and the search for a causality, and statistical learning, the primary objective of which is the prediction of an interest variable? They seem to effectively have numerous methods in common. The RFAS wishes to further investigate these links. The Revue is anticipating pedagogical articles on these technically complex issues.

In that same vein, a reflection on the evolution of the public, French, European, and even international statistical system, as it relates to the development of Big data, would be welcome.

More specifically in the pharmacology field, studies “in real life” most often allow the analysis of prescription drug abuse and undesirable side effects. There is however a difference of opinion about how valuable these studies are for demonstrating drug efficacy, though it is complicated to attribute an effect to any particular factor outside of a laboratory, unlike in double blind randomized studies. In the face of the influx of an ever-increasing amount of rich data, the possibility of going back to real life studies, at least as a supplement for clinical trials, nevertheless represents a major issue not only in terms of the costs for pharmaceutical laboratories but also and especially in terms of expertise and therefore decision-making for the public authorities.

Moreover, it is valid to question the status of these new data, through the lens used by classical statisticians: their quality, reproducibility, representativeness and the context of their production. The opposition which exists between representativeness and exhaustiveness particularly constitute one of the dead-ends in the contemporary discourse around Big data. It is not likely that these “traditional” concerns, which ensure the scientific quality of statistics, will disappear: the RFAS willingly welcomes contributions that explore this question.

Theme 2 : What Big data uses are at the core of our social protection system?

Big data radically changes entire sectors of economic and social organizations (transport, telecommunications, industry, finance, insurance...). Like many other industry and service sectors, the majority of social protection organizations are and will increasingly be impacted by Big data. They must therefore confront the unique issues they present : new services, new partnerships, efficiency, but also the weakening of existing relationships, even the undermining of the social protection model from the 20th century, which is tied to professional or national solidarity, organized around the redistribution of monetary resources that cover the principal risks of life and of services which are not often individualized.

In the field of health, the Ministry of Health and Social Affairs launched a reflection in 2015 which questioned the possible uses of Big data in health. Big data, linked to open potentialities by genomic and biotechnologies, gives rise to significant attempts in the area of prevention, patient care, precision medicine, even personalized medicine. Cancer, for example, is a pathology whose management may experience considerable progress in the future, with “customized” treatments based on the genetic and biological specificities of the tumor by taking into account the patient’s environment and way of life. The knowledge of the links and causalities between environment and health should also greatly benefit the contributions of Big data.

Beyond health, Big data could without a doubt give rise to new uses in other social protection domains, in order to better fight against job loss, poverty and social exclusion, disability... The employment and professional training sectors in particular come to mind: experiments conducted on Big data appear which aim to better assist recruiters and job-seekers. Examples and information that can be drawn from it would be welcomed.

The UN, through the *United Nations Global Pulse* initiative, as well as the World Bank, also promote the essential role that Big data may play in the fight against poverty in developing countries. Experiences in this area could be reported. In developed nations, could it not come to the aid of the socially excluded, for example in order to combat social isolation by means of social networks and connected objects?

All fields of social protection which involve services can be influenced by Big data in time, through the creation of platforms (whether they be collaborative or allow the development of two-sided markets ²) and the implementation of personalized services: companionship for disabled people (handicapped, dependent), child care, etc. These new uses however still call quality and security norms into question, along with those of their modality framework.

In each of these fields, we will investigate the role that financial institutional actors (creators, promoters, new service financiers?) and organizations (public and private) which provide services, as well as the regulatory role of the State.³

As soon as Big data brings new uses, innovative organizations or collaborations, we can investigate the capacity to improve the global efficiency of our social protection system, either by improving the quality of the services rendered, or by mastering or reducing social expenditure . We are thinking of the following examples:

- the improvement of user or patient services and the creation of new services : information on the quality of care or social protection services, personalized medicine, access to healthcare, telemedicine, platforms, in all the social protection domains, service quality management, enterprise creation and management of tax levies, etc. ;

² http://social-sante.gouv.fr/IMG/pdf/colloque_big_data_synthese.pdf.

See also the Montaigne Institute report, <http://francestrategie1727.fr/wp-content/uploads/2016/02/resume-objets-connectes-institut-montaigne.pdf>

³ These markets all bring two distinct categories of clients together through the use of an intermediary or a "platform", See for example http://www.autoritedelaconurrence.fr/doc/ocde_cp_06_2009.pdf

- the best management of resources, in the sanitation sector as well as in the social-medical sector: healthcare management, coordination of actors, implementation of best practice recommendations and the detection of risky medical practices, purchasing management, etc.;
- the evaluation of public policies, their efficiency and their impact on citizens ;
- the reduction of unused benefits by the early detection of potential beneficiaries who do not use their benefits ;
- fraud detection, whether it come from allocators, professionals, or patients;

Theoretical and prospective articles, and other more practical works that describe past or current experiences, are welcome.

The operation of social protection stakeholders can also be studied: these new tools and services may have effectively a lasting impact on existing organizations, in the realms of information systems, human resources, and relationships with the insured.

Theme 3 : How to reconcile Big data and solidarity?

Big data can favor new services and solidarities, but it can also threaten existing solidarities and increase inequality. This RFAS project would like to explore this tension which exists at the heart of these new envisioned uses. The field of investigation is vast, prospective, and it can evidently lead to a kind of alarmism if one is not careful. The objective is for this project to be pragmatic and to place itself as close as possible to the known or envisioned uses for Big data.

The most thoroughly covered problematic concerns changes in the healthcare field. Digital tools make data education necessary; the correct usage of algorithms rests on their transparency and readability. In a more general way, we often associate digital health, Big data and patient autonomy. But this is not possible if the patient is not informed, capable and concerned about their own health. Big data can thus pose a risk of growing segmentation between "autonomous" patients and others. In the same way, the development of more predictive medicine, leading to probabilistic diagnoses, can create inequalities between informed patients and others, by making decisions a matter of personal responsibility when this is not always possible. It begs the question of whether or not patients have the right to not be informed of their condition: a patient may not want to know their risk of developing one disease or another. Thus, the growth of Big data invites us to revisit the concepts of patient rights and health democracy.

Another use of Big data in the healthcare field which is often discussed has to do with insurance and the risk of coverage and contract individualization. If the provisions strictly govern the modalities of the contract segmentation according to individual characteristics, if measures exist or have been reinforced in the Jan 2016 French Public Health Law to organize, in the healthcare field, a "right to be forgotten" past a certain period to facilitate access to insurance and credit to people who have been affected by specific health problems. If the state of healthcare is itself recognized as a criterion through which discrimination can be characterized (cf. art. 225-1 of the Penal Code), Big data nevertheless puts these principles in tension and calls for a discussion of the ethical and regulatory framework. The most telling example is that of health insurance. The contradiction between the individualization that is made possible by these increasingly abundant data and the pooling of risks which is at the

very heart of the principal of insurance, is today a subject of much reflection and worry for all insurance stakeholders.

It would be interesting to report on international experiences with the modulation of private health insurance coverage pricing according to effectively observed behaviors. Does this create a consciousness on the part of insured persons or does it rather lead to the worsening of certain inequalities, notably in patient care?

Certainly, all the uses of Big data by health insurers do not directly concern the individual payer; some of the uses primarily envisioned thus far are geared towards patient support, prevention and information services. But do these uses not also potentially lead to discriminatory or exclusionary practices towards "bad risks" (meaning people for whom health care costs are higher than their contributions)?

Beyond just healthcare and insurance, other social protection sectors can see their current solidarities weakened by the new tools and practices of Big data. Questions of a very different nature can therefore be asked. If professional integration relies more and more on digital tools, what about those who cannot master them? If Big data, linked to these new domestic and surveillance tools, improves companionship for the homebound elderly population, will it be a supplement for existing formal and informal assistance or does it risk supplanting it in whole or in part, to the detriment of interpersonal time which is necessary for successful aging?

In the long run, Big data also poses the question of its impact on social inequalities, whether they be related to health, employment, working conditions, wages, or education. Can we evaluate this? What should we expect from it? "Impartial" algorithms (if such a concept even exists) that allow for a total objectification of decision-making in the allocation of housing, day care placement, school choice? Well-informed people who will know how to profit off of it and use it and exploit it and other less-informed people who may suffer it as a result?

More generally, a still rather theoretical questioning at the stage of development that Big data has reached, may be centered around the reconciliation of a digital social protection system with the 20th century model of social protection. The current apparatuses are marked by a great complexity, rely on numerous stakeholders and financing circuits which are not always visible to citizens. How can this situation be reconciled with the new principles of digital, dematerialized, personalized, adaptable and experimental services? Is the individualization of services which is at the core of Big data compatible with the national solidarity that is the foundation of social protection ? Under what conditions?

Essentially, does Big data risk crystallizing the situation that Pierre Rosanvallon announced twenty years ago, the welfare state questioned by the logic of mutualisation under the pressures of the rise of individualism? Former solidarities do not only weaken due to individualism: they exist parallel to new digital communities, potentially rich of new forms of aid, which may in time be on the same level as the traditional solidarities within the welfare state.

Theme 4 : Big data, a societal change?

Big data is certainly no Big Brother, not yet in any case. But it does have some of its same characteristics, regarding its claim to welcome and analyze a large amount of data on all aspects of human life, both on the individual and collective level. Evidently, this inherent dimension of Big data has been the subject of numerous debates and warnings. How to ensure

respect for private life and the protection of personal data? Does the old concept of a private life still mean anything or must that concept evolve?

Public authorities make an effort to ensure that people's rights are respected in terms of the conservation and processing of data. But can we say as much for all of the data that we generate and make available without necessarily being aware of it? Mobile telephone data, connected objects, emails, internet requests... All of this information that we sometimes forget about identify us and are used, unknown to us, to define new algorithms which are used most often for financial gain. The protection of private life, protected legally by the January 1978 no. 78-17 French law on computing, files and liberties and reinforced in Europe by the recent general regulation on the protection of data⁴, could become an essential issue in social protection, public or private.

Taking the results of these algorithms into account when making human decisions (for example, by "automating" management) also constitutes a new field of inquiry and challenging research. If the algorithmic rationality can lead to progress, individually and collectively (in the healthcare field especially), it also entails limits and risks that must be framed and addressed.

Could not personal support services (including for the most vulnerable) to raise awareness of their rights as regards their personal data, consent and issues linked to algorithms, constitute a new service, even a new public service?

Additional information on the content of this appeal for contributions can be obtained by contacting Magali Beffy or Franck von Lenep, who are in charge of preparing the issue, at the following addresses:

magali.beffy@sante.gouv.fr

and

rfas-drees@sante.gouv.fr

Authors wishing to propose a relevant article to the review will need to send it along with a summary and an author's introduction (cf. "advice to authors" by RFAS [online])

<http://drees.social-sante.gouv.fr/etudes-et-statistiques/publications/revue-francaise-des-affaires-sociales/>) to this address:

rfas-drees@sante.gouv.fr

Before March 31st, 2017

⁴ EU REGULATION 2016/679 of 27 April 2016 relative to the protection of physical persons with regards to the processing of personal data and the free circulation of data
<http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679>