STG Resilience Papers

The Data-Driven Pandemic:
A New Conceptualization of the Data Society

Daniel Innerarity, *Professor of Political Philosophy at the University of the Basque Country, Ikerbasque Foundation for Science, and Part-time Professor at the EUI School of Transnational Governance, Florence, Italy*

May 2021

Summary:

This first pandemic of the data society raises the disturbing question of to what extent we have the infrastructure and, above all, the right concepts to understand and combat it. We must start from the realisation that data is not neutral, and our data collection systems have blind spots. This difficulty of having good information is exacerbated by the fact that we live, as is often said, in a post-truth era. To address these difficulties, a conceptual shift is proposed away from measuring society by the notion of the majority towards a greater focus on groups and individuals.

1. Introduction

The coronavirus is the first pandemic of the data society, of the information age. The large quantity of data generated and analysed about the virus and its effects make this crisis the first "data-driven pandemic". Communications about and scrutiny of the data about infections and deaths from the coronavirus have become a daily ritual. It is no surprise that the numbers, the comparisons and the categorizations have been very important to the way governments have managed this crisis. The data publicly clarify the actual situation and guide the decisions that should be adopted. Recourse to the data allows us to re-establish a baseline of credibility that justifies the control measures and limits placed on activities. Quantification is especially seductive at times of uncertainty because it allows knowledge to be organized and simplified, making it easier to take decisions (Merry 2016; Porter 1995). Measurements give us a way to manage complexity and reduce uncertainty.

2. Data is not neutral

Despite everything, it feels like the myth that the quantity of data sufficed to grapple with reality has been refuted because the data alone did not allow us to grapple with the complexity of the phenomenon. The pandemic has revealed that our data infrastructures are insufficient for resolving social crises. First off, there is the problem of insufficient or poor-quality data. In fact, data from the pandemic are scarce, and they are fragmented by different national public health policies. Even the data about the number of deaths have been uncertain.

Not only has there been a scarcity of data, but there are also errors in the way the data are interpreted and in the very configuration of our informational spaces, which also spread extravagant disinformation. But it may be that our principal ineptitude stems, paradoxically, from some excessiveness and uncritical confidence when it comes to the existing data. From the earliest days of the pandemic, with more or less success, governments have reported on the increase in the numbers of people infected, contact tracing and hospital occupancy rates, and that information was disseminated by the means of communication on a daily basis. What has been less common is to ask about the ways in which, through that quantification, "the conditions of production influence the kinds of knowledge" (Davis / Fisher / Kingsbury 2012, 4) or the social effects that arise from quantification. My hypothesis is that "dataism" (van Dijck 2014), in other words, the belief that quantification produces truth, privileges a false sense of objectivity and provides a deceptive certainty that impedes the comprehensive knowledge of reality upon which corresponding decisions should be based.

3. Data collection systems have blind spots

In order to fight a pandemic effectively, one must know the ways in which it is propagated and the extent to which it affects different types of people. I am concerned that, when data are presented as if they were neutral, we are seduced into believing that the data are precise and we do not need to be concerned about context. To circumvent that risk, we should practice what has been called "post-Cartesian doubt" (Amoore 2019, 149) regarding the data and their technologies. In other words, we should engage in doubt about that which supposedly resolves our doubt, the doubt that must be exercised at a time when algorithms are the primary way of providing evidence. There are many biases inherent to any production, analysis and visualization of data, but the most disturbing of all of them is the assumption that the data are neutral, as if they were some kind of apolitical referees of truth. The same practices of gathering, analysis and visualization of the data lead us to ignore certain aspects of reality. We should not forget the surprise caused by the massive infection rates in certain parts of the

population, such as gig workers, prison inmates or the elderly in nursing homes. We should also remember the limited effectiveness of some of the general recommendations given precisely because different family, living or work-place realities were not taken into sufficient consideration. On the other hand, many countries contain undocumented individuals who are concerned about the encroachments on privacy that might jeopardize their ability to remain in the country or their access to health care. Also, for many people who are just scraping by on their salaries, a quarantine might be economically unsustainable, while healthcare authorities assume that people can allow themselves the luxury of staying home for two weeks.

The dominant discourse claims that our systems of analysis and surveillance are very precise, which is something that their proponents celebrate and their critics lament, but the reality may be something else entirely. One could say that "surveillance capitalism" (Zuboff 2019) is overvalued and that, instead of complete knowledge of reality, the supposedly omniscient system has revealed errors in the analysis of reality. They do not know too much about us, but too little.

The techniques that are used to obtain data make the infection rates among certain social groups invisible. We tend to forget that the measures to reduce risk reflect and encourage specific behaviours. They place people into categories and create a "standard human" who responds to a very limited and selective vision of society (Epstein 2009, 36). The groups that are least visible tend to be the ones that do not hew as closely to the norms of behaviour that are used to carry out data analysis, and these are often the people who are most likely to be infected and to be contagious. Health strategies that are meant to channel people's behaviour are ineffective, for example, with people who disobey the recommendations out of pure economic necessity or immigrants who move to areas of higher risk. It is very important to keep this in mind when the vaccines arrive, in other words, when we need to implement all the decisions that have to do with the distribution, prioritization or the corresponding communication policies regarding vaccines.

The problem with "translating social life into commensurable categories so that different events become instances of the same thing" (Merry 2016, 27) can be somewhat dangerous when there are forms of inequality or vulnerability. Some subjects are overrepresented while the system of governance fails to detect others, such as migrants without official documentation or those who work without a contract. If the data are generated by consumerism, mobility or activity on the social networks, representation will favour those who produce more data in those areas. Data collection systems have a blind eye that excludes certain people, precisely those who are most vulnerable, from

mitigation strategies. This blindness has been corrected, but perhaps only partially and too late. The approach many countries were taking to the pandemic changed when it was revealed that certain population groups (ethnic minorities, certain workers or the elderly in care facilities) had disproportionately elevated infection rates. In Great Britain, for example, they began to discuss the ethnicity of those who died, which then allowed the collection of data that were more useful when it came to carrying out preventive policies. This attention to the particular is one of the issues that remains to be addressed in our systems of quantification. The gathering of data should include people who are uninsured, who do not have residence permits or access to health services. These are the people who are most frequently infected and, therefore, the most contagious.

The uncertainty not resolved by the data is often related to people who are forgotten, sectors of society that are less visible because of their identity or work. These are often people with less personal autonomy. This has implications in the technological measures that are applied to confront the pandemic through the use of data, for example, through contact tracing apps. In the first place, contact tracing strategies tend to prioritize, generally in a non-intentional fashion, certain types of people and to marginalize others: undocumented immigrants may be afraid of being turned over to the authorities; many people will suffer through the illness in the solitude of their own homes or on the streets, and no one will gather data about them. The people who design the apps only have a certain type of user in mind, principally someone with digital skills and the financial standing to have a smart phone with an updated operating system.

These contact tracing apps have provided us with very useful information, but they have some limitations such as, for example, the false security they can give their users or the fact that they read location more than types of behaviour. The virus is not only transmitted by the means that can be detected by using the app. The contact tracing that mobile phones can realize is useful to replace the more expensive but more precise contact tracing that consists of asking people about their contacts. Furthermore, these mechanisms will only work if people trust the authorities and the centralized management of the data, which is far from being the case in a society where the pandemic has only increased the growing sense of distrust.

4. Data in the "post-truth" era

The pandemic has burst onto the stage in a world in which there is, at the same time, access to scientific knowledge, a chaotic digital news environment and distrust of experts and toward governments. This

environment presents special difficulties, including in the data themselves and their trustworthiness for managing the pandemic.

One factor that might explain our relative failure when it comes to governing this crisis is the "post-truth" attitude that has taken hold in today's social life, where objective facts seem to have less impact on our opinions, whether public or private, than appeals to emotion and personal beliefs (Shelton 2020, 1). One part of this disregard for the truth can be attributed to the actions of some governments that have hidden or manipulated data. More concerning, however, is the confusion and errors that have been produced from actual data, which have not been placed into context or analysed correctly. This reveals that the data are as conclusive as they are malleable and that anyone can present them so that they favour what one wants to say. Sanctimonious guardians of the data tend to defend them as if they were protecting us from ideologization. However, the data are not necessarily the opposite of ideological obfuscation; they may favour objectivity, but they can also be used in the service of any ideology. While this is the part of our confusion that is most crude, it is less concerning, because the most troubling aspect of this distortion of reality stems from structural causes and is not due to the deliberate intention to conceal or lie. I am referring to the ambiguous relationship that our current news environment has with the truth. There we find unprecedented access to knowledge alongside the free dissemination of errors, whether in the form of disinformation or extravagant conspiracy theories. In this "infodemic", fake news is expanding more rapidly than the virus itself (United Nations Department of Global Communications 2020).

There is a type of disinformation that is very connected to the very nature of social networks, and it contrasts with the potential that had been assigned to the networks to respond to these crises in a more effective fashion. One of the things that this pandemic places into question is the widespread belief that social networks could be systems of early surveillance to alert us to the development of illnesses. The idea is that digital footprints will make threats such as the coronavirus visible before governments or scientists do. The data that are circulating on social networks are not free of biases, and they coexist with the proliferation of fake news. Disinformation about the pandemic is due to the existence of bots—it turns out that more than half of the Twitter accounts that expressed opinions about the pandemic were bots (Hao 2020)—but it is even more concerning to confirm the number of people who participate in the spread of this false information. Disinformation has weakened citizen confidence in the authorities and has reduced the effect of the health measures that attempted to motivate prudent behaviours in the citizenry, such as wearing masks, social distancing or lockdowns.

The data-driven tendencies of recent years have contributed to this post-truth society. It leads to data without context and without a coherent narrative that could account for what was going on. Our own management of the data may be creating more confusion than comprehension. One need not have an express intention to confuse to end up making us all quite confused. It is true that journalists and sociologists have done a great job of communicating and visualizing the data of the pandemic. I am not judging their intentions but attempting to point out the unintended consequences of a certain management of the data for which we have not yet developed an appropriate culture. The redundancy of data that is afforded to us every day in maps, numbers and charts barely allows us to distinguish one data point from another (mortality versus fatality, contagion versus infection or the reasons deaths are increasing when there are less people infected). We cannot understand the meaning of what is happening. Another example of this is how the emphasis on the continuous updated representation of data can limit our perception to that which is most urgent and make us unable to comprehend the ways in which this type of crisis stems from processes that act over a longer temporal scale. In this context, it is not surprising that conspiracy theories seem more appealing.

5. Focus on the collective and specific, not the majority

The way out of this health crisis should lead us to different ways of being and knowing. This includes ways of understanding and being in the world implied by the technology for analysing data. A society constructed upon data finds it very hard to integrate into its infrastructure and governance other alternative ways of knowing and existing beyond the standard ones. Measuring and tracing have been more important for governments than understanding exactly what should be measured and traced. We must flip the terms of discussion and ask not about the data that will lead to certain policies but about the data that are needed to take the political decisions that are required.

Expanding our gaze toward those who do not tend to be the focus of attention could help us understand society from the logic of the collective and not from the notion of the majority. To understand and manage a contagious society, it is incomparably more useful to focus on the category of the collective than on the category of the majority. We need to develop a new type of attention toward social reality that is drawn to questions of the collective and specific situations. A change in the line of care also requires a change in the way of understanding the data. "Science and policy could be able to control the pandemic better by addressing the sources of uncertainty and missing data not as gaps in the information landscape, but as individuals who are likely to be members of less-visible and less powerful groups" (Taylor 2020, 1). It would mean seeing society as people and groups, not as

populations, which would allow us to take particular vulnerabilities into consideration and, therefore, take special care with those spaces of infection. We could then talk about a democracy of data, not so much from the habitual perspective that focuses on who owns the information, but concentrating instead on whether the data represent the whole society, the collective, every man and woman. This requires a different conception of the data, because it would mean not focusing on the majority but on diversity, on concrete experiences such as those of the economically disadvantaged or socially excluded. Instead of a herd governed by the statistical norm, we would have a mosaic of different vulnerabilities. The way out of the pandemic demands a change in our conception of the data and therefore a change in our way of understanding society.

6. Conclusion and recommendations

The pandemic has given us the opportunity to test the resilience of our systems for interpreting social reality, and most of the failures stem from an incorrect interpretation of the data, for which it is recommended:

1. Always bear in mind that data are not neutral and that they need appropriate interpretation tools that correct their biases and compensate for their blind spots.

2. The main conceptual revolution to which this pandemic invites us is to think society outside of standardised normality and the category of the majority. From a logic of care, it seems more appropriate to focus on those who are most likely to be infected and contagious, that is, to think of data that reflect the reality of groups and individuals.

Reference:

Amoore, Louise (2019), Doubt and the Algorithm: On the partial accounts of machine learning, *Theory, Culture & Society* 36 / 6, 147–169.

Davis, Kevin / Fisher, Angelina / Kingsbury, Benedict et al. (eds.) (2012), *Governance by Indicators. Global Power through Quantification and Rankings*, Oxford: Oxford University Press.

Epstein, Steven (2009), Beyond the Standard Human, in Lampland, Martha / Star, Susan Leigh (eds.), *Standards and Their Stories. How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca & London: Cornell University Press, 35–53.

Hao, Karen (2020), Nearly half of Twitter accounts pushing to reopen America may be bots, *MIT Technology Review*, 21 May. Available at: https://www.technologyreview.com/2020/05/21/1002105/covid-bot-twitter-accounts-push-to-reopen-america/

Merry, Sally Engle (2016), *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*, Chicago and London: University of Chicago Press.

Porter, Theodore (1995), *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*, Princeton University Press.

Shelton, Taylor (2020), A post-truth pandemic?, *Big Data & Society* 1 / 6, 1-10.

Taylor, Linnet (2020), The price of certainty: How the politics of pandemic data demand an ethics of care, *Big Data & Society* 1 / 7, 1-10.

United Nations Department of Global Communications (2020), *UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis. United Nations COVID-19 Response*. 31 March. Available at: https://www.un.org/ en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-andcybercrime-covid-19

Van Dijck, José (2014), Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology, *Surveillance & Society* 12 / 2, 197–208.

Zuboff, Shoshana (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: PublicAffairs.