UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2020/109

Data, global development, and COVID-19

Lessons and consequences

Wim Naudé[1] and Ricardo Vinuesa[2]

September 2020

Abstract: The COVID-19 pandemic holds at least seven lessons for the relationship between data-driven decision making, the use of artificial intelligence, and development. These are that (1) in a global crisis, the shifting value of data creates policy pitfalls; (2) predictions of crises and how they play out are mostly wrong and can cause unhelpful panics; (3) digital deluges are as significant a problem as lack of data; (4) data deprivation and digital divides can deepen inequalities and hamper global coordination; (5) data creates regulatory dilemmas; (6) interoperability and reuse are critical aspects of data-driven decision making but are neglected; and (7) decentralization of data gathering and data use may reduce vulnerabilities to risk, and strengthen resilience of countries and regions. The consequences for data policy and data science are that trust, equality, context, and political leadership are as important, if not more, than technology.

Key words: data science, artificial intelligence, COVID-19, development, global crisis

JEL classification: I19, O20, O32, O39

Acknowledgements: This paper builds on, and substantially extends, an earlier background note of Wim Naudé (2020a) written for UNU-WIDER, and for which he gratefully acknowledges the encouragement and comments of Kunal Sen, Lorraine Telfer-Taivainen, and Timothy Shipp, noting the usual disclaimer. Ricardo Vinuesa acknowledges support from the Swedish Research Council (VR).

[1] Technology and Innovation Management, RWTH Aachen University, Aachen, Germany; IZA Institute of Labor Economics, Bonn, Germany; MSM, Maastricht, the Netherlands; Africa Study Centre, University of Leiden, Leiden, the Netherlands, corresponding author: naude@time.rwth-aachen.de; [2] SimEx/FLOW, Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

This study has been prepared within the UNU-WIDER project on Academic excellence.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-866-5

https://doi.org/10.35188/UNU-WIDER/2020/866-5

Typescript prepared by Gary Smith.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Since the outbreak of the COVID-19 pandemic in December 2019, there has been a rush to harness data in the fight. Being a combined health and economic crisis, the pandemic is one of the most significant global development crises since the Second World War. Data, and the tools based on it (such as artificial intelligence (AI)), can help in decision making in the short-term fight on both fronts—health and economics—and more broadly in global development over the long run. For instance, in terms of reducing the health impact, data is essential for tracking and predicting the spread of the infection. It can also, as inputs into AI-based tools and scientific research, help make diagnoses and prognoses, and contribute towards the search for treatments and a vaccine (Luengo-Oroz et al. 2020). The health impact can further be mitigated, given data to inform decisions on non-pharmaceutical interventions (NPIs)—for instance, data is the essential output of contact tracing (smartphone applications, here denoted as apps) and camera surveillance, both key tools of social control that can be used to help identify and isolate those who are infected, and monitor and enforce compliance with lockdown and social distancing measures (e.g., Ferretti et al. 2020; Naudé 2020a). In terms of fighting the economic impacts of the pandemic, data is crucial to understand the relationship between health outcomes, lockdown measures, and economic activity, to trace the decline in production and consumption, to gauge the (re)distributive impact of the pandemic in terms of how it affects sectors, households, firms, and countries differently (e.g. Eichenbaum et al. 2020), and finally to help design and target financial and other assistance measures for households and firms (e.g., Bartik et al. 2020; Casado et al. 2020). In sum, the COVID-19 pandemic poses an urgent and serious global development challenge that has and will continue to place high expectations and growing demands on data and data scientists.[1]

This paper explores what lessons can so far be learned on the value of data and data-driven decision making in dealing with a global development crisis. The COVID-19 pandemic holds a number of lessons about the value, use, potential, and dangers of data-driven decision making. These lessons have stark consequences for global development, as well as for data policy in general, as they highlight the importance of non-technological aspects such as trust, equality, context, and political leadership. These lessons furthermore point to the tenuous relationship between data and information, and moreover the dangers of dealing with either too little data, too much data, or misinformation—i.e. an *infodemic*—during a crisis and more generally in tackling the world's grand development challenges. Moreover, data has side-effects: it comes inevitably with benefits as well as costs, and some of the costs are due not only to weaknesses in human institutions and governance, or to misuse of data, but to the inherent nature of data and information.

The rest of the paper proceeds as follows. In Section 2, seven lessons on data and policy are discussed. In Section 3, four consequences for global development policy and data policy are outlined. Section 4 concludes.

[1] Before the COVID-19 pandemic, the business and decision-making value of ICT (information and communication technology) and data were substantial and increasing. Production of ICT goods and services was valued in 2015 to at 6.5 per cent of world gross domestic product (Farboodi and Veldkamp 2019). The global data market was estimated to be worth US$26 billion in 2019, growing in the preceding period by more than 20 per cent per year (Chen et al. 2020). The COVID-19 pandemic has given a boost to the digitalization of the global economy and hence the global ICT and data markets.

2 Seven data lessons from the pandemic

2.1 The shifting value of data

During a global crisis, the value of data can shift significantly. The most valuable data is obviously real-time, high-frequency, reliable data. But not all real-time, high-frequency data is equally reliable or useful. And often, medium-term or recent data tends to become less valuable, and historical data more valuable. These rapid shifts in the value of data create huge pitfalls for decision making during a crisis, as they can lead to reliance on either unreliable and unverified data or useless, inaccurate data, or reliance on informed guessing, and moreover can create incentives for data manipulation and data manufacturing. As is shown in this section, all these pitfalls were encountered during the first six months of the COVID-19 pandemic.

Decision making during a crisis such as the COVID-19 pandemic requires real-time, high-frequency data to be able to track the virus, and to target responses—both health and economic assistance—accurately. Speedy reaction is of the essence. In the COVID-19 pandemic, governments have been blamed for responding either too slowly, or to have overreacted when eventually they did respond (Aksoy et al. 2020; Boretti 2020). In both cases, it may have been due partly to the problem of the shifting value of data. First, while related to the known SARS-virus family, the SARS-CoV-2 virus that causes the COVID-19 disease was essentially novel. Thus, once the disease was declared a pandemic, there was a lack of data with which to calculate certain critical parameters of epidemiological models (Leon et al. 2020), such as the reproduction number (R_0) and the case fatality risk (CFR). In the case of COVID-19, lack of reliable data on these parameters has led to reliance on unverified and biased data (e.g. taken from a small number of Chinese hospitals), and to informed guessing, and hence widely different health and economic policy responses. This has led to some confusion regarding the recommendations by health authorities on social distancing and use of protective masks,[2] which in many cases were ineffective and outdated (Asadi et al. 2020). In Section 2.2 this is further discussed in the context of panics and prediction errors.

The demand for high-frequency and real-time data during a global crisis can far outstrip the supply thereof. In fact, the crisis can make the situation worse by compromising the supply of such data. For instance, in the case of health-related data, real-time big data tend not to be available, or only available after a lag, due to the facts that (1) 'in public health, critical surveillance systems remain primarily based on manually collected and coded data, which are slow to collect and difficult to disseminate' (Callaghan 2020: 1), and (2) data-privacy regulations in health limit communication of and sharing between health centres. In the case of economic and business data, high-frequency and real-time data suffers due to the crisis leading to a direct reduction in these data flows. Ducharme et al. (2020) noted that the COVID-19 pandemic has disrupted the normal flow of essential socioeconomic data needed to track key economic indicators. Just as in the case of public health data, most economic indicators are collected by statistical agencies on a low-frequency (quarterly or annual) basis—not in time for immediate access.[3]

---

[2] Despite the recommendations by the European Centre for Disease Prevention and Control (ECDC) and the World Health Organization (WHO), certain European countries still do not recommend the use of masks for the general population. This is a critical failure of data-driven decision making—leadership—due to inadequate response in the critical light of the data on the airborne transmission of COVID-19 (Fears et al. 2020; Morawska and Milton 2020).

[3] There are, however, some positive signs that the situation is improving, with more real-time economic data sets becoming available and being used during the pandemic. Casado et al. (2020) discuss, for instance, two such data sets in the USA, namely weekly administrative data on unemployment, and daily transaction level data on economic activity. They show how this data can be used to tailor fiscal support in order to sustain consumption spending.

When the demand for certain data far outstrips the supply, it may create perverse incentives for the manufacturing of data, and for the misuse of data for the spread of misinformation and disinformation. Instead of relevant quality and high-frequency data, data deluges—too much noise[4]—can make matters worse. In Section 2.3 these are discussed in the context of digital deluges.

Another aspect of the shifting value of data which contributes to inefficiencies in the market for data is that a global crisis can result in depressing the value of pre-crisis data—for example, the data for the 5–10 years preceding the global crisis. This is not to discard such data completely—there are indeed potentially useful correlations to draw between features of a country or region just before the pandemic broke out, and the impact of the crisis, as for instance Knittel and Ozaltun (2020) illustrate. The weakness is that recent but pre-crisis data underpins models and assumptions that often have been made redundant by the nature of the crisis being experienced. This is because of two reasons: first, the crisis presents an outlier event, producing huge amounts of novel data. This reduces the usefulness of prediction models (calibrated on recent data) not only in health and medical sciences, but also in economics, finance, transport, logistics, travel, and retail—among other areas (Naudé 2020a; Rowan 2020). Second, and related to the first point, is that the data just preceding the crisis may contain little information useful for understanding the correlates and causal relationships at work in a pandemic, as it does not contain the extremes in terms of data to estimate the tail risk of the pandemic (Cirillo and Taleb 2020).

On the other hand, deep historical data can become more valuable. For instance, data on the historical impacts of pandemics—such as the 1918 flu, but also pandemics and the ways societies have dealt with them during the past 500 years—can be valuable in a number of ways for policy-making. First, as Cirillo and Taleb (2020) illustrate, historical data on epidemics and pandemics can provide better data on their tail risks. These authors use data on pandemics stretching back 2,500 years and conclude that 'the distribution of the victims of infectious diseases is extremely fat-tailed' (Cirillo and Taleb 2020: 606).

Another lesson from deep historical data is that the effects of pandemics can be very heterogeneous across regions and localities, as Carillo and Jappelli (2020) show in the case of Italy during the 1918 flu. A third lesson is that the after-effects of pandemics can be surprisingly long-lasting. Almond (2006), for example, found that the 1918 flu pandemic's effects were felt into the 1960s and 1980s through the effects on children still *in utero* at the time of the pandemic. Naudé (2020b) refers to the work of Jorda et al. (2020), who studied the macroeconomic impacts of major pandemics in Europe over the past 500 years. They found that the after-effects of such pandemics last up to 40 years, and moreover that 'pandemics are followed by sustained periods—over multiple decades—with depressed investment opportunities, possibly due to excess capital per unit of surviving labour, and/or heightened desires to save, possibly due to an increase in precautionary saving or a rebuilding of depleted wealth' (Jorda et al. 2020: 14).

2.2 Predictions and panic

The COVID-19 pandemic has shown that despite the best efforts of modellers using the available data and tools, predictions and forecasts—of both the spread of the infection and its health and economic consequences—have been far off. Indeed, prediction models have been shown to be fairly inaccurate and unreliable, despite increased computing power and bigger data sets. Tsikala Vafea et al. (2020)

[4] Acemoglu et al. (2019) illustrate the likelihood of excessive data sharing on digital online platforms as a result of externalities in the sharing of data (for instance, data from person X may be correlated with that of person Y, so that information from Y may be inferred even though Y did not share their data (see also Bergemann and Bonatti 2019)). Such excessive data sharing reduces the price of data and drives a wedge between the social value and private value of data—which in turn can reduce people's concerns about data privacy and ownership. The bigger point is that data markets are like other markets, subject to failures and inefficiencies.

noted in particular concerns about the various mathematical models (mainly SIR models—that is, those based on the labels susceptible, infectious, or recovered) used to predict how many people could be affected and what the mortality rate could be, emphasizing the lack of reliable data stating that 'Significant information about the transmission of the virus remains unknown, thus limiting the precision of forecasts' (Tsikala Vafea et al. 2020).

Wrong predictions—based on inaccurate data and incorrect assumptions—carry a cost in terms of public panic, raised anxiety, and herding behaviour. For example, the UK's decision to impose strict lockdown and social distancing measures in mid-March 2020 was strongly influenced by a paper (not peer reviewed) by Ferguson et al. (2020) from Imperial College, published on 16 March 2020 (Adam 2020). This paper also influenced the policy responses in the USA and Canada (Avery et al. 2020). It made a shocking prediction: 'in an unmitigated epidemic, we would predict approximately 510,000 deaths in GB and 2.2 million in the US, not accounting for the potential negative effects of health systems being overwhelmed on mortality' (Ferguson et al. 2020: 7). This paper has attracted heated discussion and even tabloid scrutiny,[5] with much criticism against its assumptions and predictions—for instance, by Avery et al. (2020) and Shen et al. (2020). The growing consensus is that its predictions of COVID-19 death rates were significantly overestimated (Boretti 2020). As Avery et al. (2020: 2) notes, the actual death rates due to COVID-19 a number of weeks later 'have only amounted to a fraction of those projected in the most pessimistic scenarios for the Imperial College model'.

While models based on inaccurate data will lead to wrong predictions—and possible panic and wrong policies as a result—the problem is compounded by the fact that these models tend not to suggest the optimal interventions that governments can take, nor what the socioeconomic consequences of interventions will be. To overcome these limitations there have been a number of extensions of and complements to epidemiological models since the pandemic started. For example, Eichenbaum et al. (2020) provide an extended SIR model that includes an economic cost–benefit analysis of the consequences of NPIs, Dignum et al. (2020) provide an agent-based framework within which to model the broader implications of NPIs, and O'Sullivan et al. (2020) extend the SIR model to include a spatial dimension.

Finally, even if data gathering succeeds in providing information on crucial assumptions and parameters in epidemiological models, the practicalities of data collection standards, data quality, and interpretational latitude have shown much scope for misunderstandings and confusion. Backhaus (2020) discusses a number of common pitfalls in the use of COVID-19 data. He makes the important point that attempts to compare policy efficacy between countries are often bedevilled by differences in measurements (for instance, in ascribing COVID-19 as the cause of death) and by invalid comparisons. For example, South Korea's CFR of 1.1 per cent in March has often been compared to Italy's 8.6 per cent, with the conclusion that COVID-19 is more deadly in Italy. However, as Backhaus (2020) illustrates, Italy's CFR is not comparable to that of South Korea because the underlying age distribution of the populations differs significantly—Italy has a much older population structure. Thus, from this, one should be careful about either fostering panic in Italy or uncritically adopting South Korean policies to combat the pandemic in Italy.

2.3 Digital deluges

Data can be devalued by digital and data deluges. This has been termed 'infodemics'. There are at least three dimensions to a digital deluge or infodemic. The first is data spikes, as people radically change their behaviour, including in a herd-like fashion, with this showing up in their digital footprints. What is referred to here by herd-like fashion is twofold; one is, for example, panic buying (or selling) which can lead to spikes in prices and/or supplies of certain items drying up. Think, for instance, of

[5] See, for example, https://tinyurl.com/yco7slv5.

the run on banks during the financial crisis and the run on toilet paper during the COVID-19 crisis (Paul and Chowdhury 2020). A second example of herd-like behaviour that contributes to data spikes is especially found in financial markets, where the greater uncertainty introduced by the crisis can lead market participants to use the digital connectivity 'to extract others' information, rather than to produce information themselves' (Farboodi and Veldkamp 2020: 2485). Both of these types of behaviour lead to invalidation of many forecasting and decision-making models, depending on the recent past (see the discussion in Section 2.2).

Social media and cell phone data can potentially be useful to gauge public sentiment or infer development and economic outcomes (e.g., Pestre et al. 2020; Restrepo-Estrada et al. 2018). However, during a crisis such as that of COVID-19 (and as confirmed from previous epidemics), social media can become too noisy to extract reliable inferences. An earlier example is the failure of Google Flu Trends, an attempt to identify and track the outbreak of flu epidemics through social media data, more than five years ago. This failure is dissected by Lazer et al. (2014), who describe noisy social media data as upending 'big data hubris and algorithm dynamics'.

The second dimension of a data deluge is a burden of too much new data, overwhelming the capacity to draw useful information out of the data in time for policy purposes. For instance, the number of scientific papers dealing with the pandemic has grown exponentially since March 2020. The COVID-19 Evidence Navigator documents and maps this growth, and is updated daily (Gruenwald et al. 2020). Literally thousands of new articles are published daily: on 5 June 2020, no fewer than 2,483 new publications were recorded in one day. For scientists searching for innovative new approaches to fight the disease, and policy makers looking to science for guidance, this poses the frustrating problem of the 'burden of knowledge' as described by (Jones 2009). Overcoming the burden of knowledge requires more and more teamwork as well as deeper and more specialized education—both requirements that will inevitably take time to be realized.

Another problem posed by much new novel data is that of establishing the veracity of the data. In the case of the deluge of scientific articles published since the outbreak of the pandemic, this has led to fears that not all scientific articles can be peer reviewed in a timely manner (the vast bulk so far published on COVID-19 has been on pre-publication servers such as *ArXiv*, and not peer reviewed), and that even articles with a flimsy scientific basis can slip through the overworked nets of peer reviewing. Two examples from the COVID-19 pandemic are, first, the study by Chavarria-Miró et al. (2020), which claimed that the novel coronavirus SARS-CoV-2 was first found in the sewage water of Barcelona in March 2019, whereas there is substantial genomic evidence indicating that the origin of the virus was around November 2019 in China (Van Dorp et al. 2020), and, second, the notorious (peer-reviewed) study by Mehra et al. (2020) on hydroxychloroquine for COVID-19, which had to be retracted due to irregularities in the employed data.

The third dimension of a data deluge is misinformation and disinformation. This is exacerbated by the high levels of digital connectivity in the world today—not just the volume of data. Misinformation and disinformation, including malicious activities of 'spammers and scammers' and conspiracy theorists, undermine the value of data and policies to counter the spread of the disease (Ball and Maxmen 2020; OECD 2020; Ortutay and Klepper 2020). It has even led to offline violence against certain groups (Velásquez et al. 2020).

Brennen et al. (2020: 1) found from a sample of 225 instances of misinformation about COVID-19 that 59 per cent consists of reconfiguration of data; in other words, that 'true information is spun, twisted, recontextualised, or reworked'. Online, such misinformation can rapidly spread far and wide; for example, a video claiming that COVID-19 can be prevented or cured using hot air from a hairdryer or sauna was watched hundreds of thousands of times.[6] Moderators of many COVID-19 support groups on Facebook have voiced their frustration at the extent of misinformation being posted and spread in these groups (Khalid 2020).

The impact of such misinformation can be significant. Bursztyn et al. (2020) studied the impact of two different TV shows on Fox News in the USA on the subsequent behaviour of viewers. The shows each promoted different viewpoints on the severity of the pandemic and measures to control it during the first two months following the outbreak. The authors conclude that greater exposure to the show that underplayed the severity of COVID-19 contributed to a higher number of cases and fatalities.

The OECD (2020) lists four actions that governments and online platform owners can take to limit the spread of misinformation and disinformation: encourage independent fact checkers,[7] improve the health-literacy of the general population, issue reports on the extent and nature of incidents, and keeping human moderators in the loop and not only rely on automated (AI-based) content moderation.

2.4 Data deprivations, data gaps, and digital divides

The concept of data deprivation was used by Serajuddin et al. (2015) with reference to the lack of sufficient and appropriate data by most developing countries to measure and track poverty. Data deprivation affects all countries, however. As was already mentioned, the COVID-19 pandemic has shown the inadequacy of high-frequency, immediately available data to track socioeconomic indicators and to provide accurate parameters for epidemiological models, and moreover that data collection processes were hampered by the impact of the pandemic, such as closing down statistical offices and making data collection difficult—much economic and public health data still require traditional data collection. The situation is much worse in developing countries, where pressure on fiscal resources is likely to lead to cutbacks in the budgets of statistical agencies.

In addition to these deprivations, the COVID-19 pandemic has exposed the extent to which available data—including big data underlying AI applications—suffers from gaps and biases. Knittel and Ozaltun (2020) estimate various multiple regression models to identify what data correlates with COVID-19 fatalities in the USA. They find 'deaths per 1,000 people are, on average, 1.262 higher in a county that has all African American residents compared to a county that has no African American residents' (Knittel and Ozaltun 2020: 2). As they control for a wide range of factors, such as access to health care and incomes, they conclude that the reason for this correlation is not clear and requires further attention from policy makers. One aspect that makes an evaluation of this result difficult is the existence of data gaps on minority groups—such as African Americans in the USA. Giest and Samuels (2020) point out in this regard that minority groups are often excluded from the processes that generate and collect data, including access to technology. Often, but not always, this reflects economic deprivation and discrimination, and as Iacobucci (2020) finds in the case of England and Wales, the health impacts of COVID-19 are much greater in deprived areas.

As such, even though governments may have access to 'big data', this data may not be representative of the population, and thus it may suffer from data gaps. Giest and Samuels (2020: 2) define data gaps as 'data for particular elements or social groups that are knowingly or unknowingly missing when policy is made on the basis of large datasets'. Data gaps mean that data is imperfect and may reflect the existing biases and discrimination in a society (Barocas and Selbst 2016). With data gaps, there is thus

[6] See https://tinyurl.com/rf9x434.

[7] See, for instance, the role of the International Fact-Checking Network (IFCN) in identifying and warning against COVID-19 misinformation and disinformation (www.poynter.org/coronavirusfactsalliance). At the time of writing, the IFCN has published more than 7,100 COVID-19 fact checks. The *Infodemic Observatory* evaluates around 4.7 million 'tweets' per day on COVID-19 for their reliability—see https://tinyurl.com/y5bfush6.

an 'incomplete' and potentially biased basis for policy-making,[8] which hinders inclusive responses to the pandemic and contributes to the unequal impact of the pandemic (see also Section 3.2).

One technology in the fight against COVID-19, the use of which has been halted in the USA due to underlying data gaps and biases, is infrared closed-circuit television (CCTV) cameras enabled with facial recognition technology (AI) (Ting et al. 2020). Such surveillance cameras can scan 200 persons per minute and recognize those whose body temperature exceeds 37.3 degrees.[9] In the USA, firms such as IBM have announced halting development of facial recognition technology, and Amazon indicated it would stop supplying the US police with this technology for at least a year. This follows increasing numbers of reports that found this technology being based on data biased against African Americans. For example, a study by Grother et al. (2019) in the USA found that facial recognition technology is up to 100 times more likely to misidentify an African American or Asian person than a white man. It is hence no surprise that African Americans and Asians are more distrustful of this technology.[10] It is important to highlight that this bias is not only explained by the use of biased training data sets: it is also due to inherent biases in the employed algorithms, and how the algorithms are used, even if the data is unbiased; see also footnote 8 in this regard (Cowgill and Tucker 2019; Denton et al. 2019).

It is well known that a digital divide characterizes the global economy in terms of different digital and data-based economic activities and infrastructures, as well as different digital and data analytics capabilities among countries (Hilbert 2016). Internet penetration in Africa in 2020 is around 39 per cent, compared to 88 per cent in Europe.[11] Moreover, as Ojanperä et al. (2019) and Graham et al. (2014, 2017) stress, internet connectivity is 'no panacea', and despite growing internet penetration, African countries contribute in a negligible way to global content creation in the digital economy, as measured for instance in terms of collaborative coding on GitHub, contributions to Wikipedia, and in terms of domain registrations. Hilbert (2016) has pointed to this as a particularly worrisome dimension of the digital divide, in particular as it indicates that the digital divide exists not just in terms of access to data and hard technology, but also in terms of 'the capacity to place the analytic treatment of data at the forefront of informed decision-making' (Hilbert 2016: 164). Moreover, in terms of data deprivation, Africa suffers not only from a lack of available data and analytical capacity, but also poor-quality data, a situation that has been termed a 'statistical tragedy' (Devarajan 2013).

Also within advanced economies digital divides exist and they hamper adjustment to a COVID-19 lockdown economy—for instance, in the UK around 10 per cent of households do not have access to the internet (West and Bergstrom 2020), and in the USA this number is put at 42 million people, with internet download speeds declining since the outbreak of the pandemic (Holpuch 2020). Digital capabilities are crucial for the extent to which countries are ready to cope with an economy in lockdown. Access to the internet is vital for doing any work and schooling from home, and for moving businesses online, as well as using e-commerce (WTO 2020). It is also vital for spreading public health information on the pandemic, and for developing and using digital contact-tracing apps.

It is not only general ICT and internet access that are important, but in the case of the pandemic, specific access to digital health systems in particular. Here, divides and unequal access have been noted

[8] A related concern has been that AI models (algorithms) trained on biased data will provide biased outcomes. Cowgill and Tucker (2019), however, show that this is not necessarily the case, pointing out that an algorithm trained on biased data can still reduce bias 'particularly if these data-sets contain noisy behavior that effectively act as experiments' (Cowgill and Tucker 2019: 3). The implication is that removing bias from data also may not necessarily lead to unbiased outcomes using algorithms: algorithmic bias is a challenge that can only be overcome if it is applied with a good understanding of and sensitivity to the context in which it is used, and if policy focuses not only on the inputs (data) but also the outcomes.

[9] See https://tinyurl.com/vhqzr43.

[10] See https://tinyurl.com/yysxlvhe.

[11] See www.internetworldstats.com.

as a serious obstacle to public health efforts to contain COVID-19 in the USA. Ramsetty and Adams (2020) describe the USA's initiatives to use telehealth-based care—for example, clinics providing free online healthcare platforms for consultation, and for referring patients to drive-through testing facilities if necessary. The authors describe how this system ran into a lack of reliable internet access, and how it 'quickly became apparent that the newly built telehealth systems created additional access hurdles for our free clinic patients, and we would soon learn that pockets existed within the larger population that were impacted by these barriers. As is often the case, those whose access was impeded were the most vulnerable to poor health outcomes related to COVID-19' (Ramsetty and Adams 2020: 1147).

The above discussion implies clearly that the extent to which there exist digital divides between countries could account for the different economic and health impacts of the disease. In these contexts of data deprivation, digital gaps, and digital divides, the increasing shutdowns of the internet by governments can be particularly pernicious. The #KeepItOn coalition documented 213 cases of internet shutdowns[12] by governments in 33 countries (13 in Africa) in 2019; moreover, there is a rising trend.[13]

2.5 Data dilemmas

There is much potential in using large-scale databases (big data) to help fight poverty; for instance, inferring poverty rates from AI analysis of satellite images or monitoring crop yields to predict famine (Burke and Lobell 2017; He et al. 2016). Such big data analytics can also be of help during crises, such as natural disasters. It has in this regard been used to trigger flood alarms by making use of social media data (e.g. Restrepo-Estrada et al. 2018); for earthquake early-warning systems (e.g. Asencio-Cortés et al. 2018; Yin et al. 2018); and to track volcanic eruptions (e.g. Gad et al. 2018).

The dilemma is that having more data and even better data analytical techniques, such as AI, however, does not guarantee that development outcomes will improve. In fact, both improvements in data and AI will have mixed outcomes. Vinuesa et al. (2020a) consider how AI can help the world in achieving the Sustainable Development Goals (SDGs). They conclude that the impact is mixed—there are many SDGs the achievement of which is likely to be complicated by the rise of AI and advanced data analytics. Managing this dilemma requires recognizing that data and the analytical tools with which it is used (e.g. AI) is endogenous to the development process and, as Hilbert (2016) stressed, to existing social and power relations in a country. He gives the example of Bangladesh, where 'when twenty million land records in Bangalore were digitised, creating a Big Data source aimed at benefiting 7 million small farmers from over 27,000 villages ... existing elites proved much more effective at exploiting the data provided, resulting in a perpetuation of existing inequalities' (Hilbert 2016: 156).

Once this dilemma is recognized, it is not only how data is used that matters, but also the very nature of data—and more broadly science—that matters. To illustrate this further, consider that the World Bank (2020) argues, uncontroversially, that the way that data is used may lead to adverse development outcomes. In particular, it argues that 'the misuse of data also poses significant development risks that can manifest themselves once again through public, private and civil society channels' (World Bank 2020: 7). In this view, the World Bank considers (more) data and data analytics to be good for development outcomes, except when it is 'misused'. Under 'misuses' of data, the common understanding is use of data that violates data security and privacy, the use of data for criminal purposes and for warfare, as well as the use of data to entrench the power of companies or political leaders (Vinuesa et al. 2020a). Data misuse can also include the selective use of data to manipulate and convince, for instance through

[12] According to the #KeepItOn coalition, an internet shutdown occurs when there is 'An intentional disruption of internet or electronic communications, rendering them inaccessible or effectively unusable, for a specific population or within a location, often to exert control over the flow of information.' See www.accessnow.org/cms/assets/uploads/2020/02/KeepItOn-2019-report-1.pdf.

[13] See www.accessnow.org/cms/assets/uploads/2020/02/KeepItOn-2019-report-1.pdf.

not allowing or admitting selection bias, using skewed and inaccurate data visualizations, and cherry-picking results and ignoring contrarian evidence (confirmation bias), among others[14]—as is often the case in misinformation and disinformation (see Section 2.3). There is already substantial agreement on these misuses, and the need for data governance and legislative measures, including independent fact checking and ethical governance of AI (Dignum 2019) and other analytic tools based on data (Battaglini and Rasmussen 2019).

The dilemma that exists, however, is that it is not only misuses of data that can be bad for development, but that the nature of data (even if not misused) and the related AI-based methods can have mixed development outcomes—as Vinuesa et al. (2020a) illustrate in the case of the SDGs. For instance, although extensive use of AI may lead to increased productivity and wealth, it will also raise the requirements (in terms of infrastructure and qualification) to benefit from it, thus leading to a net increase in inequalities. Essentially, the use of AI requires a global perspective in order to produce a positive impact on development: extensive work on preservation of species (SDGs 14 and 15) may have a detrimental effect on the environment (thus hindering the achievement of SDG 13). Even if the problem at hand can be narrowed down and clearly formulated, the possibilities enabled by AI and data may lead to ethical debates related to cultural differences worldwide: see, for example, the case of what ethical guidelines autonomous vehicles should follow, as studied by Awad et al. (2018) through the 'moral machine experiment'.

Finally, a significant practical data dilemma exists in using AI-based techniques in the fight against COVID-19 in developing countries. The dilemma is that countries would need complementary infrastructures and institutions, but deciding which particular kinds of institutions and infrastructures would perhaps imply that countries experiment more and gather further data, something that the need for a fast response will not allow for, nor for which sufficient fiscal leeway exists. As Blumenstock (2020) points out, 'Rigorously demonstrating that phone data can be used to predict poverty in a controlled research environment is one thing. Quickly putting this idea into operation through a complex political bureaucracy in a country of 160 million people is another. Might people with multiple phones accidentally receive multiple payouts? Will those without phones, who are presumably the most poor and vulnerable, be missed altogether? Won't some people change the way they use their phones to game the system? Right now, frankly, these questions don't have answers.'

2.6 Interoperability and reuse

Lack of interoperability and difficulty reusing data have limited the value that could be derived from the available data, including from large, open databases. One of the laudable initiatives of the COVID-19 crisis has been the efforts to create large, open data sets. A shortcoming of these data sets is that they are 'hyper-fragmented' (Luengo-Oroz et al. 2020).

Evaluating the use of these data sets in the fight against COVID-19, Alamo et al. (2020) and Luengo-Oroz et al. (2020) are critical about the lack of operability and limited use and reuse of these databases. The latter emphasized the need for interoperability, stating that 'from the epidemiological perspective, global standards and interoperability between databases could enable coordinated response and decision-making at global, national and local levels' (Luengo-Oroz et al. 2020: 296). The former concludes that 'the open datasets available presently are locally collected, imprecise with different criteria (lack of standardization on data collection), inconsistent with data models, and incomplete' (Alamo et al. 2020: 2).

An example of this is the significant differences in criteria for reporting mortality rates of COVID-19 among countries (Backhaus 2020). Such inconsistencies are clearly observed when comparing the of-

[14] West and Bergstrom (2020) argue, perhaps too optimistically, that 'While BS increasingly appears clad in the trappings of stats and data graphics, one doesn't need an advanced degree in science or mathematics to see through it.'

ficially reported figures and the excess deaths[15] (i.e. the number of deaths above the average). Leon et al. (2020) refer to these inconsistencies, pointing out that countries differ in their practices in classifying deaths as due to COVID-19 or to an underlying pre-existing condition that a person may have had. They therefore recommend, in order to make comparisons possible and evaluate the effectiveness of differing national efforts to combat the disease, that the best measure to use will be counts of weekly excess deaths, and that 'the counts would be of deaths by all causes combined, thus side-stepping issues of what is or is not a death attributable to COVID-19' (Leon et al. 2020: e81).

The use and reuse of large, open databases during the COVID-19 crisis has been, up to this point, far from optimal. In a critical review, Alamo et al. (2020) conclude that these open databases are not optimally used, nor very reusable. Access to and use of open data in the fight against COVID-19 have been compromised by the use of a variety of data formats, changing and non-uniform criteria for measurement, and continual changes in database structure and locations. Data reuse has been complicated by weaknesses such as 'lack of an API to access individual data in the data sources ... This forces the reuses to update the full dataset daily', as well as a 'lack of geolocalization contents' and little 'standardization effort' (Alamo et al. 2020: 24). Finally, as will be discussed in Section 3.1 on trust, the protection of data rights and data privacy through appropriate regulations and laws is necessary to reduce market failures for data gathering, storage, use, and reuse, and herein the current lack of international coordination of regulations on data usage and protection has been shown to lead to sub-optimal outcomes in terms of global welfare (Chen et al. 2020).

2.7 Decentralized data gathering and use

One of the valuable lessons that historical data on pandemics has offered, is that locality matters. The impact of a pandemic on health and economic outcomes tends to be very heterogeneous across regions. So, for instance, during the 1918 flu pandemic, Italian localities experienced significant differences in mortality (Carillo and Jappelli 2020). So far, much the same is being experienced with COVID-19. For example, in the UK mortality rates have differed hugely across hospitals, from 12.5 per cent to 80 per cent of persons hospitalized with the disease.[16] See also similar evidence of spatial heterogeneity for the case of Germany (Kuebart and Stabler 2020), England and Wales (Iacobucci 2020), and New Zealand (O'Sullivan et al. 2020). In the case of the USA the spatial impact has been very uneven: 72 per cent of counties did not record a single death due to COVID-19 by April 2020 (Desmet and Wacziarg 2020).

Therefore, given the need for reliable, real-time data (see Section 2.1), decentralized data gathering and use across heterogeneous communities allows more useful data to be gathered, on which for instance, NPIs can be customized (Casado et al. 2020). Generalized lockdowns of a whole country come with a huge economic cost, and constitute a sub-optimal approach given that the virus and economic impacts are spatially concentrated. Furthermore, having access to reliable decentralized data helps to evaluate what works best and what works not so well in terms of pharmaceutical and non-pharmaceutical interventions (Aubrecht et al. 2020). Hence, Erondu and Hustedt (2020) conclude that 'what we have learned from regions, states, and countries that have been more successful than others in managing the epidemic, is that the more granular and local the data, the more useful'.

Decentralized data gathering and use may also help to strengthen the trust that citizens have in their data not being misused and their privacy not being violated. One of the principles of the EU's General Data Protection Regulations (GDPR) is that of data minimization, which states that the processing of personal data needs to be 'adequate, relevant and limited to what is necessary in relation to the purposes

[15] See www.bbc.com/news/world-53073046.

[16] See https://tinyurl.com/yadll23a.

for which they are processed'.[17] The EU and the European Data Protection Board (EDPB) consider the decentralized collection and storage of personal data to be more in line with this principle than the centralization of personal data (Rossello and Dewitte 2020).

The decentralized collection and storage of data—and hence the better safeguarding of personal data—are crucial in the use of contact-tracing apps to help identify and isolate infected persons, and in essence to manage a 'smart lockdown' (Eichenbaum et al. 2020). Countries such as Hong Kong, Singapore, and South Korea have been lauded for their use of contact-tracing apps in reducing the spread of the pandemic (Huang et al. 2020). As a result, and again in an effort to limit the use of broad lockdowns, an increasing number of countries have been developing contact-tracing apps. At least 25 countries had, already by mid-April 2020, resorted to contact-tracing technologies (Gershgorn 2020). If the use of a contact-tracing app is voluntary, then the public will be more likely to trust the system being able to guarantee data privacy, which is essential to getting enough people to use the app, which is a pre-condition for its success (Ferretti et al. 2020). In this sense, the centralized approach initially proposed by the UK for their contact-tracing app (NHS 2020) had to be abandoned and replaced by a decentralized approach. The importance of public trust in the government strategy for digital contact tracing is highlighted by Vinuesa et al. (2020b) and is further discussed in Section 3.1.

At the time of writing, at least four privacy-preserving decentralized methods or protocols for contact tracing have been proposed,[18] namely DP-3T, TCN Coalition, PACT (MIT), and PACT (UW) methods. The DP-3T (Decentralized Privacy-Preserving Proximity Tracing) method uses the Bluetooth function of a smartphone to link anonymously with other smartphones in proximity, and stores any data collected on the users' smartphones, not in a central cloud. Note, however, that even apps based on Bluetooth technology still have shortcomings in the context of COVID-19, as O'Neill (2020) discusses, including the possibility of signalling too many false positives and requiring algorithmic adjustment of signals.

3 Implications for data and development policy and practice

The seven lessons discussed in the previous section imply that in a world that is more interdependent and vulnerable to global hazards, the rise in the prominence of data will have both positive and negative impacts on development. Four related aspects that should be seen as complementary factors in the value of data are trust, equality, context, and political leadership. This section will expound on these four aspects.

3.1 Trust

Fukuyama (2020) stressed the importance of trust in government as a critical ingredient of a country's resilience in the face of a risk and disaster such as COVID-19. The World Bank (2020: 12) is therefore correct in recognizing that

> Harnessing the full development potential of data entails its repeated reuse to extract a wide range of different insights. This in turn rests on a transaction between the data provider and the data user that is founded on trust. Without adequate 'safeguards', the provider may lack the confidence that data can be shared without potential abuses. Such 'safeguards' include data protection regulations, including the right of consent on the part of the data provider and a series of obligations on the part of the data user.

[17] See GDPR, Article 5.1c at https://gdpr-info.eu/art-5-gdpr.

[18] See https://drive.google.com/file/d/1OQg2dxPu-x-RZzETlpV3lFa259Nrpk1J/view.

One of the areas where there has perhaps been most debate over the gathering and use of data during the COVID-19 pandemic has been in connection with the deployment and use of contact-tracing apps—as was outlined in Section 2.7. These apps can enable contact tracing to be conducted faster and more effectively. Ferretti et al. (2020) argue that the spread of the virus is too fast for manual contact tracing, and that digital contact tracing through smartphone apps is an alternative to lockdowns in the fight against COVID-19. Note that this is only effective if a significant fraction of the population uses the app. There are good and bad ways of using this technology as far as the data is concerned. In Section 2.7 it was indicated that privacy-preserving contact tracing is better from a data security and safeguarding point of view, and thus better at inducing trust. This is the approach that many EU countries have been following.

In contrast, China has adopted an approach that is not based on data privacy or trust. It has deployed an app that is required for citizens to move between sectors and to use public transportation (Mozur et al. 2020). It is alarming that China collects data from citizen mobility and diagnostics centrally, and through AI-based analysis the citizens are issued an indicator (from red to green) which then enforces restrictions regarding their movement. This has raised concerns (since such extensive surveillance sets a precedent and it may stay in place after the COVID-19 pandemic), and there is an ongoing discussion regarding the ethical requirements of such digital contact-tracing approaches (EDPB 2020; Mello and Wang 2020; Morley et al. 2020). A comprehensive framework was developed by Vinuesa et al. (2020b) to assess the compliance of such apps in terms of rights of citizens, technology, and governance.

Whether governments should be able to gather this information and precisely how it should be used, stored, and even destroyed has been the topic of many debates. On 19 April 2020, a large number of scientists published a 'Joint Statement on Contact Tracing', stressing the central importance of trust in this matter, stating that it is 'crucial that citizens trust the applications in order to produce sufficient uptake to make a difference in tackling the crisis. It is vital that, in coming out of the current crisis, we do not create a tool that enables large scale data collection on the population, either now or at a later time.'[19]

The danger that the COVID-19 pandemic has illustrated clearly in this regard is that citizens in many countries do not trust their governments sufficiently, and with reason. For instance, participation online is in most African countries heavily controlled, under state surveillance, and circumscribed, according to the Freedom on the Net report.[20] In other countries, such as China, state surveillance is even worse—and reflects deep distrust between citizens and the state. Feldstein (2019) notes that 75 countries globally are 'actively' using AI surveillance technologies, and a growing number are sourcing these technologies from China.

Citizens in most countries do not see safeguards on their data and data privacy as either being in place or being sufficient. Outside of the EU with its GDPR[21]—probably the most advanced data rights protection legislation in the world—or California's Consumer Privacy Act in the USA, few countries have sufficient legislative protection for data rights. Chen et al. (2020: 5) estimate that around 42 per cent of countries 'still do not have legislation or regulation on data usage and protection'. This includes most countries in sub-Saharan Africa, the world's poorest region. The Malabo Convention on Cybersecurity and Data Rights, an African Union (AU) initiative to regulate data ownership and usage in Africa, has only been ratified by eight member states.[22]

---

[19] From the Joint Statement on Contact Tracing, 19 April 2020. Available at: https://giuper.github.io/JointStatement.pdf.

[20] See https://tinyurl.com/y4qsexuf.

[21] See the European Union General Data Protection Regulation (2016). Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[22] See https://tinyurl.com/y5mggp6m.

3.2 Equality and inclusion

What is defining the debate on data and global development is the presence of significant digital divides, as was discussed in Section 2.4. In an economy in which data is a critical resource for value creation, and fuelling creative destruction as well as posing new threats to work, incomes, and livelihoods, the presence of digital divides bodes ill for growing inequality (Hilbert 2016), as well as for growing market domination—concerns that have, among others, led to fears of 'platform capitalism' (Srnicek 2016).

Naudé (2020b) surveys the various channels through which the COVID-19 crisis will exacerbate inequality within and between countries, and between generations. The rise of the digital and data-based global economy and the manner in which data-driven decision making and new digital technologies can be used for smart lockdowns and to ameliorate the impacts of lockdowns is an important channel. For instance, large digital platform firms that benefit from data-driven network economies will be bigger and more dominant in their markets, and the decline in small businesses (due to bankruptcies) and new start-ups would leave fewer possible competitors. In this respect, historical data can be very insightful in the present pandemic. For one, in all previous large pandemics (where more than 100,000 people died) inequality increased (Furceri et al. 2020).

Second, the growing dominance of a few already powerful firms has also been documented in past pandemics. In the fourteenth century, for instance, the Black Death boosted the market dominance and wealth of a few well-positioned incumbents. As Russell and Parker (2020) describe, after the Black Death market concentration increased and the influence of big business on government grew, an outcome also likely in the present crisis. This would suggest that post-COVID, the taxing and regulating of global digital platforms (through, for instance, a digital service tax as proposed by the EU), as well as the promotion of efforts to strengthen the ownership rights of consumers over their own data (see Jones and Tonetti 2020), may become more acute.

Two further ways in which the data-driven economy will be a channel through which the COVID-19 pandemic will exacerbate inequality and exclusion is due to the differences in how easily people can work from home, and how susceptible their job is to automation. Not all countries and regions have the ICT infrastructure to allow working from home or moving a business online (Brynjolfsson et al. 2020), and the pandemic will accelerate current trends towards digitization and automation (Bloom and Prettner 2020; Schrage 2020).

The above ways in which the data-driven economy will exacerbate inequality will in particular affect young people and those in developing countries. This is because the impacts of lockdowns accrue disproportionately to young people, and those in developing countries. Note that in developing countries the percentage of young people is disproportionately high. Young people in developing countries will not only lose their jobs and future career prospects to a greater extent than older people, but also they will lose out in terms of time lost in education. The latter problem will be more serious for countries and young people without access to ICT to allow for online studying.

3.3 Context: no one-size-fits-all approach

The need for decentralized approaches to data gathering and acting on data, and the communication of reliable data so as to influence behaviour, is a more optimal strategy than centralization due to the proximity of local authorities to their populations. Local authorities therefore matter when there is heterogeneity across populations, and the need for local context to be taken into account.

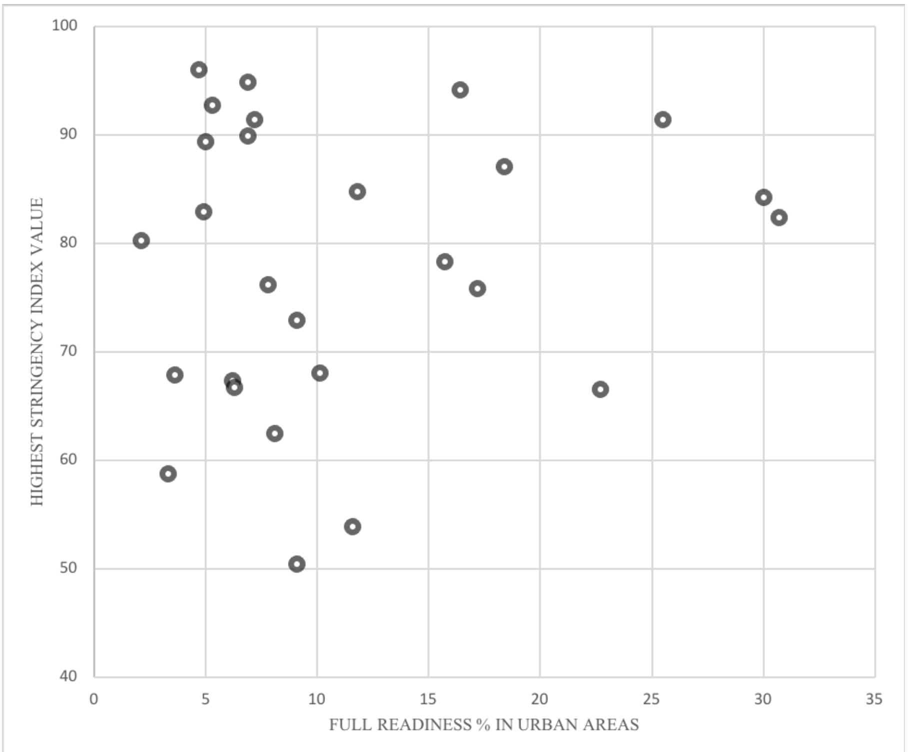
This applies within countries, and also globally between countries. There is therefore no one-size-fits-all approach to development policy. While this wisdom was made clear by the general failure of standardized structural adjustment programmes based on the so-called Washington Consensus (e.g. Nayyar

2016), this costly obtained wisdom was clearly forgotten during the pandemic, with countries all rushing to impose the same Western-style measures, despite huge differences in context, and despite fundamental uncertainties due to data deprivation. For instance, Erondu and Hustedt (2020) bemoaned the fact that

> mimicking of Western measures to combat COVID-19 ... has sometimes resulted in more dangerous outcomes for already impoverished and struggling populations ... In Zimbabwe, a small country experiencing the second highest inflation worldwide and 90 percent unemployment, lockdowns were extended indefinitely when the country reached 46 reported COVID-19 cases with four attributable deaths.

In Figure 1, the relationship is shown between the readiness of a country for typical lockdown measures and the stringency of measures adopted in 30 sub-Saharan African countries, based on data from Egger et al. (2020). Countries with the lowest lockdown readiness, such as Sierra Leone, Uganda, and Zimbabwe, were among those implementing the most stringent lockdowns.

Figure 1: Ready or not: no significant relationship between lockdown readiness and stringency of lockdown measures in sub-Saharan Africa



Source: authors' compilation based on data from Egger et al. (2020) on lockdown readiness, and the Oxford COVID-19 Government Response Tracker (OxCGRT) on lockdown stringency.

The consequences for development of policy choices that ignore data on the local context are that 'severe economic deprivation among those not prepared for the lockdowns may lead to non-compliance with lockdown and possibly a backlash against distrusted institutions, risking social unrest' (Egger et al. 2020: 10).

It is therefore important to consider not only the risks associated with the spread of the pandemic, but also those connected to raising unrest and reducing trust, which further undermine the potential of a region or locality to generate information and reliable data for decision making. The reason, as explained by Hidalgo (2015) and based on Fukuyama (1995), is that societies with higher levels of trust find it easier to establish and develop the social networks that result in access to localized information. By ignoring the local context (local data), policy makers raise distrust, which in turn reduces the ability of the local context to provide useful information to members of society, a fact that makes policy makers susceptible to turning to externally derived solutions. Thus, a vicious circle can ensue.

3.4 Political leadership

There are, in particular, four areas where political leadership vis-à-vis data and a global crisis has shown to be important during the COVID-19 pandemic. The first is in the speed and flexibility of response. Overall, democratic leadership is superior to autocratic approaches. Although autocratic regimes can act more swiftly and decisively, such regimes are less flexible for learning, experimenting, correcting errors, and being open and transparent in sharing data (Aksoy et al. 2020; Frey et al. 2020; Naudé 2020b). Not all democracies, however, responded equally quickly or decisively. Aksoy et al. (2020) found that countries where public attention regarding the pandemic was high, as reflected in Google searches on the topic, were more proactive. Using data on Google searches in 174 countries between 1 January and 31 March 2020, and data on institutional quality from the World Bank's Worldwide Governance Indicators, these authors found that in countries with both high public attention and high institutional quality, the governments' implementation of NPI was generally much faster—examples include Denmark, Ireland, South Korea, and Switzerland. Because public attention differs between countries and also within countries, this reinforces the point about local context made in section 3.3—local political leadership engaging with and understanding local public attention, conditional on this public attention being able to be freely expressed via the internet, can make a significant difference.

In this respect, an important note should be made regarding openness of debate and public attention: the controversial approach taken by the Swedish Public Health Agency (Giesecke 2020) received significant criticism from a large group of Swedish experts (King et al. 2020). The Swedish media has been criticized for fiercely attacking the points of view that were in disagreement with that of the Public Health Agency, which threatens the quality (and even the existence) of much-needed debates during a global crisis.[23] Relatedly, in the USA it has been reported that a data scientist in Florida was fired for not manipulating COVID-19 data on a public dashboard,[24] and that the Indonesian president suppressed COVID-19 data in his country.[25] Suppression of debates and data about COVID-19 is unfortunately more common than these examples, and a stark reminder that political leadership most often fails during global crises in terms of openness and transparency.

It is essential to develop adequate procedures and platforms for data communication, and for critical and open discussion of the data. For instance, Battiston and Gamba (2020) find, based on data from Italy, that municipalities with large numbers of initial infections had a subsequent lower reproduction number (R_0) and that this is due to people adhering more strictly to lockdown and social distancing measures. Hence, they conclude that 'This result calls for a transparent, real-time distribution of detailed epidemiological data, as such data affects the behaviour of populations in areas affected by the outbreak' (Battiston and Gamba 2020: 1). In this respect, good political leadership is essential to build and maintain the societal trust that is needed to fight the pandemic—as the discussion in Section 3.1 emphasized.

A second area where political leadership has shown to be particularly important is in global and national (intra-state) cooperation. Getting value from data in facing a global crisis, as well as tackling global development challenges, requires far better global cooperation—as well as cooperation between various regional governments—than what we have so far seen during the COVID-19 pandemic. In Section 2.7, the importance of decentralized data gathering and use was discussed. To get the most value from such decentralized data gathering and use, high levels of effective and efficient cooperation between

[23] See https://tinyurl.com/y3uol6ws.

[24] See https://tinyurl.com/yxkyzfpy.

[25] See https://tinyurl.com/uahew5z.

different subnational or decentralized government leadership bodies are needed. Iverson and Barbier (2020) provide a model showing that in countries where subnational governments coordinate their lockdown policies, there is better control of the pandemic, as compared to countries where the subnational governments act unilaterally.

The lack of global coordination on data sharing and cooperation to find a vaccine also reflects that 'the pandemic is stoking xenophobia, hate and exclusion, posing a far-reaching—and potentially long-lasting—threat to human rights' (Luengo-Oroz et al. 2020: 296). This lack of global cooperation reflects seriously on global political leadership. It also touches on the role and responsibilities of global organizations and how they handle geo-political tensions. In this respect, the role of the WHO has come under fire—as well as praise—during the pandemic (as did that of global organizations such as the World Bank and the International Monetary Fund during previous crises). As noted by Fidler (2020), 'Today, in the COVID-19 pandemic, the world's most powerful countries are demanding that WHO follow their respective sovereign interests for reasons that have little to do with global health ... The manner in which China and the United States politicized COVID-19 for geopolitical purposes bodes ill for international health cooperation.' Thus, on the one hand the WHO is criticized for bowing to pressure from China, and on the other hand it is threatened with financial sanctions by the USA.

The lack of global cooperation on health issues as reflected in the controversies surrounding the WHO will undoubtedly worsen the unequal impact of the pandemic, and worsen the global outcomes for all countries, but for developing countries in particular. At this point in time, developing countries should prioritize their scarce resources to prop up their health sectors and provide social security to their citizens. In essence, they should not be investing their resources in developing data-driven AI analytics in the hope of improving hospital efficiencies, or of finding a vaccine. Rather, at this juncture the most optimal global outcomes will require synergistic international cooperation: developing countries should not invest in research to fight COVID-19, and a reliable international network should be in place to share data and results. The threat is that, in the past, developed countries have had better access to vaccines: this cannot be repeated (Fidler 2010). For example, there are ongoing efforts in South America to develop alternative vaccines due to the fear of not having access to a new vaccine (Mega 2020), but those resources would be better used for testing, implementing protective measures such as masks and social distancing, and some sort of controlled, digital contact tracing. Developing countries should also partake in the gathering and building of large public databases on which to train AI. The costs of doing so are small, and the potential benefits, given the need for unbiased and representative data on the pandemic, are great. The construction of such databases should be seen as an investment against future pandemics.

A third area or requirement of leadership is to avoid mission creep such as would be the case with the use of privacy-invading contact-tracing apps and other surveillance technologies (e.g. cameras), as well as the overriding of constitutional freedoms through regulations restricting personal freedom, choice, and expression. The danger is, therefore, as Harari (2020) warned, that once the outbreak is over, erosion of data privacy would not be reversed, and governments would continue to keep intrusive tabs on their populations. They can even potentially use the data obtained in the fight against COVID-19 for other, nefarious, purposes. It is therefore important to decentralize data gathering and use, as well as set strict limitations on the time that governments can keep the data. The earlier expressed intention of Public Health England to keep for 20 years data gathered through its test-and-trace programme seems excessive (Hern 2020).

Finally, good leadership, in combination with an understanding of the information contained in historical data, would contribute to better preparedness for global disasters on a country and international level.

The COVID-19 pandemic was not, as some have stated, a 'black swan'.[26] There were plenty of deep historical precedents, as well as a growing understanding of the likelihood of a global pandemic. For example, on 21 November 2017, writing in *Foreign Affairs*, Ingelsby and Haas (2017) warned that 'the potential remains for a lethal strain of influenza or other contagious pathogen to overwhelm global health care systems by spreading at a rate that outpaces our ability to respond. In such a calamitous scenario, neither the United States nor other countries would be well enough equipped to contain it, increasing the potential for a true national or global catastrophe.'

4 Concluding remarks

This paper has discussed seven major lessons from the COVID-19 pandemic for the relationship between data-driven decision making and development. Four consequences for data policy were drawn out. These seven major lessons are: (1) in a global crisis, the shifting value of data creates policy pitfalls; (2) forecasts of crises and how they play out need to be viewed with caution, and can cause unhelpful panics; (3) digital deluges can be just as significant a problem as too little data; (4) data deprivation and digital divides can deepen inequalities during a crisis and hamper global coordination; (5) data creates regulatory dilemmas since not all the negative consequences of more data are due to misuse only; (6) interoperability and reuse are two aspects of data-driven decision making that are critical but neglected; and (7) more decentralization of data gathering and data use is required to reduce vulnerabilities to risk and strengthen resilience of countries and regions.

These lessons have at least four consequences for data policy—namely that trust, equality, context, and political leadership are vital foundations for data policy. The difficulty taking these lessons and their consequences on board is that, as the COVID-19 pandemic illustrated, despite the prevalence of massive amounts of (big) data, critical information for decision making (in the case of COVID-19, such as required by epidemiological models) will either not be available in time, or will be skewed by use of unreliable, biased data. Decision making in the modern, interconnected, and data-rich world is therefore subject to fundamental uncertainty. What does this uncertainty imply for leadership, trust, equality, and the local context? To draw these things out, there are interesting parallels with one of the grand global challenges of the time: climate change. Just as epidemiological models lacked accurate information at the start of the pandemic, key models of climate change—*integrated assessment models*—face a similar lack of accurate information and are hence subject to fundamental uncertainty. According to Heal (2017: 1047), 'How to describe the uncertainties we face, how to model them, and what constitutes rational choice in the light of them, are therefore issues that are central to climate policy analysis.'

One may extend this statement to global development in general. Facing systemic uncertainty, policy choices and debates may better prioritize risk aversion and risk management—in other words, development policy should aim more to avoid low-probability but highly catastrophic events from occurring, if at all possible, and to improve resilience once they do occur. These are the types of events with 'fat tails' as explained by Cirillo and Taleb (2020). Avoidance of, preparedness for, and resilience in the face of such potentially catastrophic fat-tail events may depend, as in Kremer's *O-Ring Theory*, on how strong the weakest link in global society is (Kremer 1993). According to Malcolm Gladwell, the COVID-19 pandemic has exposed global society as a 'complex weak-link society.'[27]

In fact, recurring future pandemics are very likely due to the reduction of biodiversity associated with human expansion (Gibb et al. 2020). Therefore, it is essential to learn from the experience accumulated

[26] See also https://tinyurl.com/yac77gtf.

[27] See https://brandondonnelly.com/2020/04/12/malcolm-gladwell-on-the-world-after-covid19.

over the past months, identify the successes and the mistakes, as well as the weak links, and develop adequate protocols for the future. The decisions taken by the governments to fight COVID-19 should be independently investigated, as proposed by García-Basteiro et al. (2020) with the aim of ensuring the use of the best possible strategies in future pandemics. Data-driven decision making, in particular by reducing structural weak links such as human cognitive biases, data gaps and disparities, and digital divides, and the related inequalities these are often associated with, is an essential pillar of resilience in global development.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2019). 'Too Much Data: Prices and Inefficiencies in Data Markets'. NBER Working Paper 26296. Cambridge, MA: NBER. https://doi.org/10.3386/w26296

Adam, D. (2020). 'Modelling the Impact: The Simulations Driving the World's Response to COVID-19'. *Nature*, 580. https://doi.org/10.1038/d41586-020-01003-6

Aksoy, C.G., B. Eichengreen, and O. Saka (2020). 'The Political Scar of Epidemics'. NBER Working Paper 2740. Cambridge, MA: NBER. https://doi.org/10.3386/w27401

Alamo, T., D. Reina, M. Mammarella, and A. Abella (2020). 'Covid-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic'. *Electronics*, 9(827). https://doi.org/10.3390/electronics9050827

Almond, D. (2006). 'Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population'. *Journal of Political Economy*, 114(4): 672–712. https://doi.org/10.1086/507154

Asadi, S., N. Bouvier, A. Wexler, and W. Ristenpart (2020). 'The Coronavirus Pandemic and Aerosols: Does COVID-19 Transmit Via Expiratory Particles?' *Aerosol Science and Technology*, 54(6): 635–38. https://doi.org/10.1080/02786826.2020.1749229

Asencio-Cortés, G., A. Morales-Esteban, X. Shang, and F. Martínez-Álvarez (2018). 'Earthquake Prediction in California Using Regression Algorithms and Cloud-Based Big Data Infrastructure'. *Computers & Geosciences*, 115: 198–210. https://doi.org/10.1016/j.cageo.2017.10.011

Aubrecht, P., J. Essink, M. Kovac, and A. Vandenberghe (2020). 'Centralized and Decentralized Responses to COVID-19 in Federal Systems: US and EU Comparisons. Law & Economics of Covid-19 Working Paper 04/2020. Rotterdam: Erasmus University Rotterdam and University of Ljubljana. https://doi.org/10.2139/ssrn.3584182

Avery, C., W. Bossert, A. Clark, G. Ellison, and S. Ellison (2020). 'Policy Implications of Models of the Spread of Coronavirus: Perspectives and Opportunities for Economists'. NBER Working Paper 27007. Cambridge, MA: NBER. https://doi.org/10.3386/w27007

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J. Bonnefon, and I. Rahwan (2018). 'The Moral Machine Experiment'. *Nature*, 563: 59–64. https://doi.org/10.1038/s41586-018-0637-6

Backhaus, A. (2020). 'Common Pitfalls in the Interpretation of COVID-19 Data and Statistics'. *Intereconomics*, 55: 162–66. https://doi.org/10.1007/s10272-020-0893-1

Ball, P., and A. Maxmen (2020). 'The Epic Battle Against Coronavirus Misinformation and Conspiracy Theories'. *Nature*, 581(7809: 371–374. https://doi.org/10.1038/d41586-020-01452-z

Barocas, S., and A. Selbst (2016). 'Big Data's Disparate Impact'. *California Law Review*, 104: 671–732. https://doi.org/10.2139/ssrn.2477899

Bartik, W., Z. Cullen, E. Glaeser, M. Luca, C. Stanton, and A. Sunderam (2020). 'The Targeting and Impact of Paycheck Protection Program Loans to Small Businesses'. NBER Working Paper 27623. Cambridge, MA: NBER. https://doi.org/10.3386/w27623

Battaglini, M., and S. Rasmussen (2019). 'Transparency, Automated Decision-Making Processes and Personal Profiling'. *Journal of Data and Privacy*, 2: 331–41.

Battiston, P., and S. Gamba (2020). 'Covid-19: R0 Is Lower Where Outbreak Is Larger'. Center for European Studies Paper 438. Cambridge, MA: Center for European Studies.

Bergemann, D., and A. Bonatti (2019). 'The Economics of Social Data: An Introduction'. Cowles Foundation Discussion Paper 2171R. New Haven, CT: Yale University. https://doi.org/10.2139/ssrn.3459793

Bloom, D., and K. Prettner (2020). 'The Macroeconomic Effects of Automation and the Role of COVID-19 in Reinforcing Their Dynamics'. VOX CEPR Policy Portal. London: CEPR.

Blumenstock, J. (2020). 'Machine Learning Can Help Get COVID-19 Aid to Those Who Need It Most'. *Nature*. https://doi.org/10.1038/d41586-020-01393-7

Boretti, A. (2020). 'After Less than 2 Months, the Simulations that Drove the World to Strict Lockdown Appear to be Wrong, the Same of the Policies they Generated'. *Health Services Research and Managerial Epidemiology*. https://doi.org/10.1177/2333392820932324

Brennen, J., F. Simon, P. Howard, and R. Nielsen (2020). 'Types, Sources, and Claims of COVID-19 Misinformation'. Oxford: Reuters Institute for the Study of Journalism, Oxford Martin Programme on Misinformation, Science and Media. University of Oxford. Available at: https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation (accessed 28 August 2020).

Brynjolfsson, E., J. Horton, A. Ozimek, D. Rock, G. Sharma, and H.-Y. TuYe (2020). 'COVID-19 and Remote Work: An Early Look at US Data'. NBER Working Paper 27344. Cambridge, MA: NBER. https://doi.org/10.3386/w27344

Burke, M., and D. Lobell (2017). 'Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems'. *Proceedings of the National Academy of Sciences*, 114(9): 2189–94. https://doi.org/10.1073/pnas.1616919114

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2020). 'Misinformation during a Pandemic'. NBER Working Paper 27417. Cambridge, MA: NBER. https://doi.org/10.3386/w27417

Callaghan, S. (2020). 'COVID-19 Is a Data Science Issue'. *Patterns*, 1: 1–3. https://doi.org/10.1016/j.patter.2020.100022

Carillo, M., and T. Jappelli (2020). 'Pandemics and Local Economic Growth: Evidence from the Great Influenza in Italy'. CEPR Discussion Paper 14849. London: CEPR.

Casado, M., B. Glennon, J. Lane, D. McQuown, D. Rich, and B. Weinberg (2020). 'The Effect of Fiscal Stimulus: Evidence from COVID-19'. NBER Working Paper 27576. Cambridge, MA: NBER. https://doi.org/10.3386/w27576

Chavarria-Miró, G., E. Anfruns-Estrada, S. Guix, M. Paraira, B. Galofré, G. Saánchez, R. Pinto, and A. Bosch (2020). 'Sentinel Surveillance of SARS-CoV-2 in Wastewater Anticipates the Occurrence of COVID-19 Cases'. *medRxiv*. https://doi.org/10.1101/2020.06.13.20129627

Chen, Y., X. Hua, and K. Maskus (2020). 'International Protection of Consumer Data'. CESifo Working Paper 8391. Munich: CESifo.

Cirillo, P., and N. Taleb (2020). 'Tail Risk of Contagious Diseases'. *Nature Physics*, 16: 606–13. https://doi.org/10.1038/s41567-020-0921-x

Cowgill, B., and C. Tucker (2019). 'Economics, Fairness and Algorithmic Bias'. *Journal of Economic Perspectives*. https://doi.org/10.2139/ssrn.3361280

Denton, E., B. Hutchinson, M. Mitchell, and T. Gebru (2019). 'Detecting Bias with Generative Counterfactual Face Attribute Augmentation'. *Preprint arXiv*, 1906.06439.

Desmet, K., and R. Wacziarg (2020). 'Understanding Spatial Variation in COVID-19 Across the United States'. NBER Working Paper 27329. Cambridge, MA: NBER. https://doi.org/10.3386/w27329

Devarajan, S. (2013). 'Africa's Statistical Tragedy'. *Review of Income and Wealth*, 59: S9–S15. https://doi.org/10.1111/roiw.12013

Dignum, F., V. Dignum, P. Davidsson, et al. (2020). 'Analysing the Combined Health, Social and Economic Impacts of the Coronavirus Pandemic Using Agent-Based Social Simulation'. *Minds & Machines*, 30: 177–94. https://doi.org/10.1007/s11023-020-09527-6

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. New York: Springer. https://doi.org/10.1007/978-3-030-30371-6

Ducharme, L., J. Tebrake, and Z. Zhan (2020). 'Keeping Economic Data Flowing during COVID-19'. *IMF Blogs*, 26 May.

EDPB (2020). 'Guidelines 04/2020 on the Use of Location Data and Contact Tracing Tools in the Context of the COVID-19 Outbreak'. Brussels: European Data Protection Board.

Egger, E.-M., S. Jones, P. Justino, I. Manhique, and R. Santos (2020). 'Africa's Lockdown Dilemma: High Poverty and Low Trust'. UNU-WIDER Working Paper 2020/76. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/833-7

Eichenbaum, M., S. Rebelo, and M. Trabandt (2020). 'The Macroeconomics of Epidemics'. NBER Working Paper 26882. Cambridge, MA: NBER. https://doi.org/10.3386/w26882

Erondu, N., and J. Hustedt (2020). 'COVID-19 Policies Not Backed by Data Do More Harm than Good'. *The New Humanitarian*, 18 June. Available at: www.thenewhumanitarian.org/opinion/2020/06/18/COVID-19-policy-data-economy-health (accessed 24 August 2020).

Farboodi, M., and L. Veldkamp (2019). 'A Growth Model of the Data Economy'. Mimeo.

Farboodi, M., and L. Veldkamp (2020). 'Long-Run Growth of Financial Data Technology'. *American Economic Review*, 110(8): 2485–523. https://doi.org/10.1257/aer.20171349

Fears, A., W. Klimstra, P. Duprex, et al. (2020). 'Persistence of Severe Acute Respiratory Syndrome Coronavirus 2 in Aerosol Suspensions'. *Emerging Infectious Diseases*, 26(9). https://doi.org/10.3201/eid2609.201806

Feldstein, S. (2019). 'The Global Expansion of AI Surveillance'. Working Paper. Washington, DC: Carnegie Endowment for International Peace.

Ferguson, N., D. Laydon, G. Nedjati-Gilani, et al. (2020). 'Report 9: Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID- 19 Mortality and Healthcare Demand'. Report. London: Imperial College London.

Ferretti, L., C. Wymant, M. Kendall, L. Zha, L. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser (2020). 'Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing'. *Science*, 368(6491). https://doi.org/10.1126/science.abb6936

Fidler, D. (2010). 'Negotiating Equitable Access to Influenza Vaccines: Global Health Diplomacy and the Controversies Surrounding Avian Influenza H5N1 and Pandemic Influenza H1N1'. *PLoS Medicine*, 1(7): e1000247. https://doi.org/10.1371/journal.pmed.1000247

Fidler, D. (2020). 'The World Health Organization and Pandemic Politics'. Think Global Health, 10 April. Available at: https://www.thinkglobalhealth.org/article/world-health-organization-and-pandemic-politics (accessed 24 August 2020).

Frey, C., G. Presidente, and C. Chen (2020). 'Covid-19 and the Future of Democracy'. VOX CEPR Policy Portal. London: CEPR.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. New York: Simon & Schuster.

Fukuyama, F. (2020). 'The Thing that Determines a Country's Resistance to the Coronavirus'. *The Atlantic*, 30 March.

Furceri, D., P. Loungani, J.D. Ostry, and P. Pizzuto (2020). 'COVID-19 Will Raise Inequality if Past Pandemics Are a Guide'. VOX CEPR Policy Portal. London: CEPR.

Gad, M., M. Elshehaly, D. Gracanin, and H. Elmongui (2018). 'A Tracking Analyst for Large 3D Spatiotemporal Data from Multiple Sources (Case Study: Tracking Volcanic Eruptions in the Atmosphere).' *Computers & Geosciences*, 111: 283–93. https://doi.org/10.1016/j.cageo.2017.10.003

García-Basteiro, A., C. Alvarez-Dardet, A. Arenas, et al (2020). 'The Need for an Independent Evaluation of the COVID-19 Response in Spain'. *The Lancet*, 396(10250): 529–30. https://doi.org/10.1016/S0140-6736(20)31713-X

Gershgorn, D. (2020). 'We Mapped How the Coronavirus Is Driving New Surveillance Programs Around the World'. Medium: OneZero, 13 April. Available at: https://onezero.medium.com/the-pandemic-is-a-trojan-horse-for-surveillance-programs-around-the-world-887fa6f12ec9 (accessed 24 August 2020).

Gibb, R., D. Redding, K. Chin, A. Donnelly, T. Blackburn, T. Newbold, and K. Jones (2020). 'Zoonotic Host Diversity Increases in Human-Dominated Ecosystems'. *Nature*, 584: 398–400. https://doi.org/10.1038/s41586-020-2562-8

Giesecke, J. (2020). 'The Invisible Pandemic'. *The Lancet*, 395(10238): E98. https://doi.org/10.1016/S0140-6736(20)31035-7

Giest, S., and A. Samuels (2020). "'For Good Measure": Data Gaps in a Big Data World'. *Policy Sciences*, 53: 559–69. https://doi.org/10.1007/s11077-020-09384-1

Graham, M., B. Hogan, R. Straumann, and A. Medhat (2014). 'Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty'. *Annals of the Association of American Geographers*, 104(4): 746–64. https://doi.org/10.1080/00045608.2014.910087

Graham, M., S. Ojanperä, M. Anwar, and N. Friederici (2017). 'Digital Connectivity and African Knowledge Economies'. *Questions de Communication*, 32: 345–60. https://doi.org/10.4000/questionsdecommunication.11579

Grother, P., M. Ngan, and K. Hanaoka (2019). 'Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects'. Report 8280. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280

Gruenwald, E., D. Antons, and T. Salge (2020). 'COVID-19 Evidence Navigator'. Aachen: Institute for Technology and Innovation Management, RWTH Aachen University.

Harari, Y. (2020). 'The World after Coronavirus'. *Financial Times*, 20 March.

He, X., E. Lau, P. Wu, et al. (2016). 'Combining Satellite Imagery and Machine Learning to Predict Poverty'. *Science*, 353: 790–94. https://doi.org/10.1126/science.aaf7894

Heal, G. (2017). 'The Economics of the Climate'. *Journal of Economic Literature*, 55(3): 1046–63. https://doi.org/10.1257/jel.20151335

Hern, A. (2020). 'Public Health England Will Keep Personal Data of People with Coronavirus for 20 Years'. *Guardian*, 28 May.

Hidalgo, C. (2015). *Why Information Grows*. London: Penguin.

Hilbert, M. (2016). 'Big Data for Development: A Review of Promises and Challenges'. *Development Policy Review*, 34(1): 135–74. https://doi.org/10.1111/dpr.12142

Holpuch, A. (2020). 'US's Digital Divide "Is Going to Kill People" as Covid-19 Exposes Inequalities'. *Guardian*, 13 April.

Huang, Y., M. Sun, and Y. Sui (2020). 'How Digital Contact Tracing Slowed Covid-19 in East Asia'. *Harvard Business Review*, 15 April.

Iacobucci, G. (2020). 'Covid-19: Deprived Areas Have the Highest Death Rates in England and Wales'. *British Medical Journal*, 369: m1810. https://doi.org/10.1136/bmj.m1810

Ingelsby, T., and B. Haas (2017). 'Ready for a Global Pandemic? The Trump Administration May Be Woefully Underprepared'. *Foreign Affairs*, 21 November.

Iverson, T., and E. Barbier (2020). 'National and Sub-National Social Distancing Responses to COVID-19'. CESifo Working Paper 8452. Munich: CESifo.

Jones, B. (2009). 'The Burden of Knowledge and the Death of Renaissance Man: Is Innovation Getting Harder?'. *Review of Economic Studies*, 76(1): 283–317. https://doi.org/10.1111/j.1467-937X.2008.00531.x

Jones, C. and C. Tonetti (2020). 'Nonrivalry and the Economics of Data'. NBER Working Paper 26260. Cambridge, MA: NBER. https://doi.org/10.3386/w26260

Jorda, O., S. Singh, and M. Taylor (2020). 'Longer-Run Economic Consequences of Pandemics'. *Covid Economics*, 1. Available at: https://cepr.org/content/covid-economics-vetted-and-real-time-papers-0 (accessed 24 August 2020).

Khalid, A. (2020). 'Moderators of Covid-19 Survivor Groups Say Keeping Up with Misinformation Is a Nightmare'. Medium: One Zero, 23 July. Available at: https://onezero.medium.com/moderators-of-covid-19-survivor-groups-say-keeping-up-with-misinformation-is-a-nightmare-6ad0d9d4b30c (accessed 24 August 2020).

King, C., L. Einhorn, N. Brusselaers, et al. (2020). 'COVID-19: A Very Visible Pandemic'. *The Lancet*, 396(10248): E95. https://doi.org/10.1016/S0140-6736(20)31672-X

Knittel, C., and B. Ozaltun (2020). 'What Does and Does Not Correlate with COVID-19 Death Rates'. NBER Working Paper 27391. Cambridge, MA: NBER. https://doi.org/10.3386/w27391

Kremer, M. (1993). 'The O-Ring Theory of Economic Development'. *Quarterly Journal of Economics*, (108): 551–75. https://doi.org/10.2307/2118400

Kuebart, A., and M. Stabler (2020). 'Infectious Diseases as Socio-spatial Processes: The COVID-19 Outbreak in Germany'. *Journal of Economic and Social Geography*. https://doi.org/10.1111/tesg.12429

Lazer, D., R. Kennedy, G. King, and A. Vespignani (2014). 'The Parable of Google Flu: Traps in Big Data Analysis'. *Science*, 343(6176): 1203–05. https://doi.org/10.1126/science.1248506

Leon, D., V. Shkolnikov, L. Smeeth, P. Magnus, M. Pechholdova, and C. Jarvis (2020). 'COVID-19: A Need for Real-Time Monitoring of Weekly Excess Deaths'. *The Lancet*, 395: e81. https://doi.org/10.1016/S0140-6736(20)30933-8

Luengo-Oroz, M., P. Hoffmann, J. Bullock, et al. (2020). 'Artificial Intelligence Cooperation to Support the Global Response to COVID-19'. *Nature Machine Intelligence*, 2: 295–97. https://doi.org/10.1038/s42256-020-0184-3

Mega, E. (2020). 'Latin American Scientists Join the Coronavirus Vaccine Race: "No One's Coming to Rescue Us"'. *Nature*, 582. https://doi.org/10.1038/d41586-020-01756-0

Mehra, M., S. Desa, F. Ruschitzka, and A. Patel (2020). 'Hydroxychloroquine or Chloroquine with or without a Macrolide for Treatment of COVID-19: A Multinational Registry Analysis'. *The Lancet*. RETRACTED. https://doi.org/10.1016/S0140-6736(20)31180-6

Mello, M., and C. Wang (2020). 'Ethics and Governance for Digital Disease Surveillance'. *Science*, 368(6494): 951–54. https://doi.org/10.1126/science.abb9045

Morawska, L., and D. Milton (2020). 'It Is Time to Address Airborne Transmission of COVID-19'. *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciaa939

Morley, J., J. Cowls, M. Taddeo, and L. Floridi (2020). 'Ethical Guidelines for COVID-19 Tracing Apps'. *Nature*, 582: 29–31. https://doi.org/10.1038/d41586-020-01578-0

Mozur, P., R. Zhong, and A. Krolik (2020). 'In Coronavirus Fight, China Gives Citizens a Color Code, with Red Flags'. *New York Times*, 1 March.

Naudé, W. (2020a). ‘Artificial Intelligence, COVID-19, and Developing Countries: Priorities and Trade-Offs’. UNU-WIDER Background Note. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/WBN/2020-4

Naudé, W. (2020b). ‘Entrepreneurial Recovery from COVID-19: Decentralization, Democratization, Demand, Distribution, and Demography’. IZA Discussion Paper 13436. Bonn: IZA.

Nayyar, D. (2016). ‘Structural Transformation in the World Economy: On the Significance of Developing Countries’. UNU-WIDER Working Paper 2016/102. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2016/146-8

NHS (2020). ‘COVID-19’. London: National Health Service, UK. Available at: https://covid19.nhs.uk (accessed 24 August 2020).

OECD (2020). ‘Start-Ups in the Time of COVID-19: Facing the Challenges, Seizing the Opportunities’. OECD Policy Responses to Coronavirus. Paris: OECD.

Ojanperä, S., M. Graham, and M. Zook (2019). ‘The Digital Knowledge Economy Index: Mapping Content Production’. *Journal of Development Studies*, 55(12): 2626–43. https://doi.org/10.1080/00220388.2018.1554208

O’Neill, P. (2020). ‘Bluetooth Contact Tracing Needs Bigger, Better Data’. *MIT Technology Review*, 22 April.

Ortutay, B., and D. Klepper (2020). ‘Virus Outbreak Means (Mis)Information Overload: How to Cope’. *AP News*, 22 March.

O’Sullivan, D., M. Gahegan, D.J. Exeter, and B. Adams (2020). ‘Spatially Explicit Models for Exploring COVID-19 Lockdown Strategies’. *Transactions in GIS*, 24(4): 967–1000. https://doi.org/10.1111/tgis.12660

Paul, S., and P. Chowdhury (2020). ‘Strategies for Managing the Impacts of Disruptions during COVID-19: An Example of Toilet Paper’. *Global Journal of Flexible Systems Management*, 21: 283–93. https://doi.org/10.1007/s40171-020-00248-4

Pestre, G., Letouzé, E., and E. Zagheni (2020). ‘The ABCDE of Big Data: Assessing Biases in Call-Detail Records for Development Estimates’. *World Bank Economic Review*, 34(S1): S89–S97. https://doi.org/10.1093/wber/lhz039

Ramsetty, A., and C. Adams (2020). ‘Impact of the Digital Divide in the Age of COVID-19’. *Journal of the American Medical Informatics Association*, 27(7): 1147–48. https://doi.org/10.1093/jamia/ocaa078

Restrepo-Estrada, C., S. Andrade, N. Abe, M. Fava, E. Mendiondo, and J. Albuquerque (2018). ‘Geo-social Media as a Proxy for Hydrometeorological Data for Streamflow Estimation and to Improve Flood Monitoring’. *Computers & Geosciences*, 111: 148–58. https://doi.org/10.1016/j.cageo.2017.10.010

Rossello, S., and P. Dewitte (2020). ‘Anonymization by Decentralization: The Case of Covid-19 Contact-Tracing Apps’. *European Law Blog*, 25 May.

Rowan, I. (2020). ‘What Happens to AI when the World Stops (COVID-19)?’ Medium: Towards Data Science, 31 March. Available at: https://towardsdatascience.com/what-happens-to-ai-when-the-world-stops-covid-19-cf905a331b2f (accessed 24 August 2020)

Russell, E., and M. Parker (2020). ‘How Pandemics Past and Present Fuel the Rise of Mega-Corporations’. *The Conversation*, 3 June.

Schrage, M. (2020). ‘Data, Not Digitalization, Transforms the Post-Pandemic Supply Chain’. *MIT Sloan Management Review*, 29 July.

Serajuddin, U., H. Uematsu, C. Wieser, N. Yoshida, and A. Dabalen (2015). ‘Data Deprivation: Another Deprivation to End’. Policy Research Working Paper 7252. Washington, DC: World Bank Group. https://doi.org/10.1596/1813-9450-7252

Shen, C., N. Taleb, and Y. Bar-Yam (2020). ‘Review of Ferguson et al “Impact of Non-Pharmaceutical Interventions..."’. New England Complex Systems Institute, 17 March. Available at: https://necsi.edu/review-of-ferguson-et-al-impact-of-non-pharmaceutical-interventions (accessed 24 August 2020).

Srnicek, N. (2016). *Platform Capitalism*. London: Polity.

Ting, D., L. Carin, V. Dzau, and T. Wong (2020). ‘Digital Technology and COVID-19’. *Nature Medicine*, 26: 459–61. https://doi.org/10.1038/s41591-020-0824-5

Tsikala Vafea, M., E. Atalla, K. Georgakas, F. Shehadeh, E.K. Mylona, M. Kalligeros, and E. Mylonakis (2020). ‘Emerging Technologies for Use in the Study, Diagnosis, and Treatment of Patients with COVID-19’. *Cellular and Molecular Bioengineering*. https://doi.org/10.1007/s12195-020-00629-w

Van Dorp, L., M. Acman, D. Richard, et al. (2020). ‘Emergence of Genomic Diversity and Recurrent Mutations in SARS-CoV-2’. *Infections, Genetics and Evolution*, 83: 104351. https://doi.org/10.1016/j.meegid.2020.104351

Velásquez, N., R. Leahy, N. Johnson Restrepo, Y. Lupu, R. Sear, N. Gabriel, O. Jha, B. Goldberg, and N. Johnson (2020). ‘Hate Multiverse Spreads Malicious COVID-19 Content Online Beyond Individual Platform Control’. *ArXiv*, 2004.00673.

Vinuesa, R., H. Azizpour, I. Leite, et al. (2020a). ‘The Role of Artificial Intelligence in Achieving the Sustainable Development Goals’. *Nature Communications*, 11(233). https://doi.org/10.1038/s41467-019-14108-y

Vinuesa, R., A. Theodorou, M. Battaglini, and V. Dignum (2020b). ‘A Socio-technical Framework for Digital Contact Tracing’. *Results in Engineering*. https://doi.org/10.1016/j.rineng.2020.100163

West, J., and C. Bergstrom (2020). ‘Hydroxychloroquine for COVID-19 Prevention? How to Separate Science from Partisanship’. *Think*, 5 August.

World Bank (2020). ‘World Development Report 2021: Data for Better Lives’. Concept Note. Washington, DC: World Bank.

WTO (2020). ‘E-Commerce, Trade and the COVID-19 Pandemic’. Information Note. Geneva: World Trade Organization.

Yin, L., J. Andrews, and T. Heaton (2018). ‘Reducing Process Delays for Real-Time Earthquake Parameter Estimation: An Application of KD Tree to Large Databases for Earthquake Early Warning’. *Computers & Geosciences*, 114: 22–29. https://doi.org/10.1016/j.cageo.2018.01.001