


# Data deprivations, data gaps and digital divides: Lessons from the COVID-19 pandemic

Big Data & Society  
July–December: 1–12  
© UNU-WIDER 2021  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20539517211025545  
journals.sagepub.com/home/bds  


Wim Naudé<sup>1,2,3</sup>  and Ricardo Vinuesa<sup>4</sup> 

## Abstract

This paper draws lessons from the COVID-19 pandemic for the relationship between data-driven decision making and global development. The lessons are that (i) users should keep in mind the shifting value of data during a crisis, and the pitfalls its use can create; (ii) predictions carry costs in terms of inertia, overreaction and herding behaviour; (iii) data can be devalued by digital and data deluges; (iv) lack of interoperability and difficulty reusing data will limit value from data; (v) data deprivation, digital gaps and digital divides are not just a by-product of unequal global development, but will magnify the unequal impacts of a global crisis, and will be magnified in turn by global crises; (vi) having more data and even better data analytical techniques, such as artificial intelligence, does not guarantee that development outcomes will improve; (vii) decentralised data gathering and use can help to build trust – particularly important for coordination of behaviour.

## Keywords

Data science, artificial intelligence, COVID-19, developing countries, global impact

## Introduction

Data-driven decision making has become increasingly important to inform policies that aim to improve global development outcomes. The United Nations' sustainable development goals (SDGs) for instance explicitly link to quantifiable metrics to track progress, and the Paris climate agreement is based on data-intensive science and modelling to set targets for reducing greenhouse gas emissions. The quest for evidence-based policymaking has led to the widespread use of randomised clinical trials to generate data on what works and what does not work in terms of development policy. And the accumulation of big data, following advances in computing power and connectivity, has led to the creation of major data-analytic units in, or aligned to, global development organisations, such as the UN Global Pulse and the Independent Evaluation Office of the Global Environmental Facility, to mention but two examples.

The COVID-19 pandemic, the most urgent global development crisis since the Second World War, has and will continue to place high expectations and growing demands on data and data scientists – not least because of the growth of big data and the many various types and

modalities of data, each coming with its own risks and challenges. The global data market, already worth an estimated US\$26 billion in 2019 has been given a boost by the COVID-19 pandemic (Chen et al., 2020). How this will subsequently affect decision making for global development is a relevant question, particularly so since even before the COVID-19 pandemic, and despite the recognition that more and better data can help in crafting development policies, it was clear that data deprivations, data gaps and digital divides were structurally contributing to development failures, for instance in perpetuating inequality, poverty and vulnerabilities (Hilbert, 2016; World Bank, 2020).

<sup>1</sup>Cork University Business School, University College Cork, Cork, Ireland

<sup>2</sup>RWTH Aachen University, Aachen, Germany

<sup>3</sup>IZA Institute of Labor Economics, Bonn, Germany

<sup>4</sup>SimEx/FLOW, Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

## Corresponding author:

Wim Naudé, Cork University Business School, University College Cork, Cork, Ireland.  
Email: wnaude@ucc.ie



In this paper, we argue that the COVID-19 pandemic has accentuated these structural problems in data deficiencies underlying development outcomes, and that it offers potentially valuable lessons to address these, post-pandemic. The rest of the paper is organised around, first, lessons for the management of a global (health) crisis (second section), and second, lessons for global development policy more generally (third section). Although we present the lessons from the COVID-19 pandemic for the relationship between data and global development separately as lessons for managing a global health crisis, and as lessons for more generally addressing global development challenges, there is no watertight distinction: global health crises (and other extreme events) and global development in general are closely intertwined. For example, resilience against health crises is needed to maintain development gains, and development gains are required to ensure better resilience against external shocks and extreme events. The fourth section concludes with a summary and conclusions for global development policy and further research.

## Lessons for managing a global (health) crisis

### *The shifting value of data*

During a global crisis, the value of data can shift rapidly. Data that underpins models and assumptions can be made redundant by the nature of a crisis for two reasons: first, a crisis such as COVID-19 is 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, 2020; 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).

The value of data can also shift in terms of becoming compromised as a result of the fact that the demand for high-frequency and real-time data during a global crisis can outstrip the supply thereof. 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 dependence on manually collected and coded data (Callaghan, 2020) and data-privacy regulations in health limit communication of and sharing between health centres. In some instances, the supply of data will be reduced – for instance because statistical

agencies have to do with reduced budgetary resources to conduct regular surveys (Ducharme et al., 2020).

When the demand for certain data 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<sup>1</sup> – can make matters worse. In section ‘Digital deluges,’ these are discussed in the context of digital deluges.

Shifting value of data as described here create pitfalls for decision making, in particular leading to (i) reliance on either unreliable and unverified data or useless, inaccurate data and models, and (ii) data manipulation and data manufacturing.

*Keeping in mind the shifting value of data and the pitfalls it can create, is a first lesson that the COVID-19 pandemic taught.*

### **Predictions, inertia, panic**

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). As a result, predictions of the extent and impact of the pandemic was, in its early stages, subject to great uncertainty (Tsikala Vafea et al., 2020). 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 different policy responses.

For example, many governments have been blamed for responding either too slowly, or to have overreacted when eventually they did respond (Aksoy et al., 2020; Boretti, 2020). A case in point is the UK, where the government was relatively slow to respond but then imposed strict lockdown and social distancing measures in mid-March 2020 following the publication a paper (then not peer reviewed) by Ferguson et al. (2020) from Imperial College (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 discussion and criticism – for instance, by Avery et al. (2020) and Shen et al. (2020). With the benefit of hindsight, it is now agreed 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.’

Parameter uncertainty and the predictions this results in can not only lead to inertia or panicked responses, but also herding behaviour. For instance, once countries where the pandemic first spread most virulently, such as China, Italy and UK imposed strict lockdown measures, almost all countries (as the *Oxford COVID-19 Government Response Tracker* documents<sup>2</sup>) subsequently rushed to impose the same Chinese and Western-style measures, despite huge differences in context, and despite fundamental uncertainties due to data deprivation. For instance, Erondy 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 world-wide and 90 percent unemployment, lockdowns were extended indefinitely when the country reached 46 reported COVID-19 cases with four attributable deaths.

In confirmation of this, Egger et al. (2020) provided an index for the readiness of a country for typical lockdown measures and the stringency of measures adopted in 30 sub-Saharan African countries. They found that countries with the lowest lockdown readiness, such as Sierra Leone, Uganda and Zimbabwe, were among those implementing the most stringent lockdowns. They noted with some concern from this that 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).

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 pitfalls in the use of COVID-19 data. He makes the point that attempts to compare policy efficacy between countries are often bedeviled 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% in March has often been compared to Italy’s 8.6%, with the conclusion that COVID-19 is more deadly in Italy. However, as

Backhaus (2020) illustrates, Italy’s case fatality risk is not comparable to that of South Korea because the underlying age distribution of the populations differ 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.

*The second lesson that the COVID-19 pandemic has taught is that predictions carry costs in terms of inertia, overreaction, and herding behaviour.*

### Digital deluges

*A third lesson taught by the COVID-19 pandemic is that data can be devalued by digital and data deluges that accompanies a global crisis.* This devaluation of data by the creation of huge volumes of data due to people’s response to the pandemic has even been given a name: ‘infodemics.’

There are at least three dimensions<sup>3</sup> 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 the run on banks during the 2009 global 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 found especially in financial markets, where the uncertainty introduced by the crisis can lead market participants to use their digital connectivity ‘to extract others’ information, rather than to produce information themselves’ (Farboodi and Veldkamp, 2020: 2485). Both of these behaviours will invalidate of forecasting and decision-making models (see also the discussion in section ‘Predictions, inertia, panic’). For example, 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, social media can become too noisy to extract reliable inferences (Lazer et al., 2014).

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 significantly since March 2020. The *COVID-19 Evidence Navigator* documents and maps this growth (Gruenwald et al., 2020). Literally thousands of new articles are published daily: on 5 June 2020, no fewer

than 2483 new publications were recorded in one day.<sup>4</sup> 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). Part of the burden of knowledge problem in the case of a pandemic is in 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 published on COVID-19 has been on pre-publication servers such as *ArXiv*), and that even articles with a flimsy scientific basis can slip through the overworked nets of peer reviewing.<sup>5</sup> Overcoming the burden of knowledge requires more and more teamwork as well as deeper and more specialised education – requirements that will inevitably take time to be realised.

The third dimension of a data deluge is misinformation and disinformation, including malicious activities of ‘spammers and scammers’ and conspiracy theorists. This undermines the value of data and policies to fight the disease (Ball and Maxmen, 2020; Ortutay and Klepper, 2020). It has 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% consists of reconfiguration of data, i.e. that ‘true information is spun, twisted, re-contextualised 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.<sup>6</sup> 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. Bursztyl 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.

### *Interoperability and reuse*

Of the many laudable initiatives undertaken during the COVID-19 crisis one has to mention efforts to create large, open data sets, for example on GitHub – see Naudé (2020) for a discussion. However, at least until the time of writing, these data sets suffered from the

shortcoming of being ‘hyper-fragmented’ (Luengo-Oroz et al., 2020).

Thus, *a fourth lesson that the COVID-19 pandemic has taught is that lack of interoperability and difficulty reusing data limits the value from data.*

The hyper-fragmentation of data sets, and hence their lack of interoperability, as well as the limited use and reuse in the fight against COVID-19, are discussed by Alamo et al. (2020) and Luengo-Oroz et al. (2020). The latter emphasised 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 standardisation on data collection), inconsistent with data models, and incomplete’ (Alamo et al., 2020: 2). An example of the lack of standardisation on data collection during the COVID-19 pandemic is the differences in criteria for reporting mortality rates<sup>7</sup> of COVID-19 among countries (Backhaus, 2020; Leon et al., 2020).

The use and reuse of large, open databases during the COVID-19 crisis has not been optimal (Alamo et al., 2020). 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 users to update the full dataset daily’, as well as a ‘lack of geolocalization contents’ and little ‘standardization effort’ (Alamo et al., 2020: 24).

Finally, the protection of data rights and data privacy 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).

### **Lessons for global development policy**

In the previous section, four lessons for the management of a global health crisis from the perspective of data were discussed. These lessons are relevant from the perspective of global development policy, given that a global health crisis is, as COVID-19 amply demonstrated, also a development crisis. Hence, it will in addition to utilising data more optimally from a health perspective, also require utilising data more optimally for development policy, as for instance reflected in

measures to eradicate poverty, reduce inequality and promote sustainable development, including climate action. In this section, we elaborate this statement by deriving three further lessons from the COVID-19 pandemic, specifically for such global development policymaking.

### *Data deprivations, data gaps, data absences 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 mentioned, the COVID-19 pandemic has shown the inadequacy of high frequency, immediately available data to track socio-economic indicators, 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 artificial intelligence (AI) applications – suffers from gaps, absences and biases. Knittel and Ozaltun (2020) use multiple regression models to identify the correlates of COVID-19 fatalities in the USA. They find ‘deaths per 1000 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. Moreover, understanding the correlation, finding causal mechanisms, and remedying these, require having more data on minority groups. 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 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). To remedy data gaps, requires bearing in mind ‘data absences’ (Leszczynski and Zook, 2020), which reflects the absence of power of marginalised groups in society. Pelizza (2020) discusses the dangers of data gaps and data absences in contributing to the spread of ‘pseudoscientific accounts’ and ‘fake news’ about minorities’ immunity to COVID-19 in the USA. With data gaps and the absence of data, policymaking may not be inclusive and may result exacerbate the unequal impact of the pandemic. As Taylor (2020: 5) stresses, ‘If policy responses around the world cannot take account of the vulnerability of groups or response systems, policymakers are blinded to the true course of the pandemic and cannot combat it effectively. If the virus lives on amongst the poor and marginalized, everyone on earth is at risk.’

These arguments are more generally applicable to global development policymaking and moreover imply that the potential reduction in spending on collecting data, as a result of the pandemic diverting resources, will disproportionately affect people in poorer countries. It was of course already well-known before the pandemic that there exists a global digital divide. 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, many poor countries, particularly those in Africa, suffers not only from a lack of available data and analytical capacity, but also poor-quality data, a situation described as a ‘statistical tragedy’ (Devarajan, 2013).

But digital divides also exist in advanced economies – for instance, in the UK around 10% of households do not have access to the internet (Watts, 2020). 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. With digital divides also existing within advanced economies, the implication is that COVID-19 may also exacerbate within-country inequality, marginalisation and exclusion.

It is not only general information and communication technology 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 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 health care 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 that data deprivation, digital gaps and digital divides within and between countries could account for the different economic and health impacts of the disease. These different impacts could, moreover, further entrench inequalities.<sup>8</sup> One way how this can happen is that the move to online business will strengthen the dominance of large digital platform firms.<sup>9</sup> A second way in which the responses to the pandemic could worsen inequality is through the differences in how easily people can work from home, and how susceptible their job is to automation (Brynjolfsson et al., 2020). A third way is that the impacts of lockdowns accrue disproportionately to young people, and those in developing countries (Greenstone and Nigam, 2020).

In sum, *a fifth lesson taught by the COVID-19 pandemic is that data deprivation, digital gaps, and digital divides are not just a by-product of unequal global development but will magnify the unequal impacts of any global crisis and will be magnified in turn by global crises.*

### 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).

However, and perhaps counter-intuitively, *a sixth lesson taught by the COVID-19 pandemic 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, improvements in data and AI will inevitably have mixed outcomes. Vinuesa et al. (2020a) consider how AI can help the world achieve the 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 recognising 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 digitized, 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 recognised, it is not only *how* data is used, that matters, but also the very *nature* of data – and more broadly science – that matters. 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 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.'

### **Decentralised data gathering and use**

Trust is a critical form of social capital that is associated with better global development outcomes (Algan and Cahuc, 2010). The World Bank (2020: 12) is therefore correct in recognising that in order to harness '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.'

*The seventh lesson that the COVID-19 pandemic has taught is that decentralised data gathering, and use, can help to build trust – and moreover that this is particularly important where citizens have low levels of trust in their governments.*

There are two reasons why decentralised data gathering and use can help build trust, also in general development policies, as the COVID-19 crisis illustrated. The first is that development outcomes and development obstacles are, similar to the impacts of external crises, spatially heterogeneous. For example, in the UK mortality rates have differed hugely across hospitals, from 12.5% to 80% of persons hospitalised with the disease (Campbell and McIntyre, 2020). 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% of counties did not record a single death due to COVID-19 by April 2020 (Desmet and Wacziarg, 2020).

As such, decentralising data gathering and use across heterogeneous communities will allow more pertinent data to be gathered, on which for instance, development policies, or in the case of a pandemic, health interventions can be customised (Casado et al., 2020). The need for decentralised 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 centralisation due to the proximity of local authorities to their populations. 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.

Local authorities therefore matter when there is heterogeneity across populations, and the need for local context to be taken into account. Of course, to get the most value from such decentralised data gathering and use, high levels of effective and efficient cooperation between different subnational or decentralised 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.

Having access to reliable decentralised data and subnational coordination and cooperation in using this data, helps to evaluate what works best and what works not so well and to allow for lesson-sharing, not only in terms of pharmaceutical and non-pharmaceutical interventions (Aubrecht et al., 2020) but also in terms of development policy (Duflo, 2017). Hence, Erundu 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.' This lesson is also, in our view, applicable to development policy more generally: by more successfully managing a challenge, whether a pandemic or absolute poverty, trust in government will be strengthened.

The second reason why decentralised data gathering, and use, can help build trust in general development policies, is that decentralised data gathering may be more privacy-preserving than centralised gathering and use – a position which the European Data Protection Board endorses.<sup>10</sup> Citizens may be more willing to share their data, if their privacy concerns are addressed (Rossello and Dewitte, 2020).

In the COVID-19 pandemic this issue was most clearly illustrated in connection with the deployment and use of contact-tracing apps, which are only effective if a significant fraction of the population uses them, which in turn will largely depend on whether the app will protect their privacy (Ferretti et al., 2020). As an example of the importance of this, the centralised approach initially proposed by the UK for their contact-tracing app (NHS, 2020) had to be abandoned and replaced by a decentralised approach.<sup>11</sup>

That the success of data-intensive technologies would benefit from decentralised gathering and use rather than top-down, centralised approaches due to being able to garner more trust was also emphasised by Sandvik (2020: 7) in the case of the Norwegian 'Smittestopp' app, and by Naudé and Cameron (2021) in the case of South Africa. A wide range of

data modalities have been combined to develop innovative strategies aimed at containing COVID-19 spread. The contact-tracing apps mentioned above combine geo-spatial data and health information, and in some contexts, these have been used to train AI algorithms. For instance, China deployed an app which was necessary to move among regions and to access e.g. the public transportation. This app would use a color code to enforce stronger or milder movement restrictions, based on an AI model that was fed with the gathered data (Mozur et al., 2020). This highlights the risks associated to the misuse of data, in particular in situations of crisis. There is a balance between the interests of the governments (to contain the pandemic) and those of the citizens (to preserve their rights), and a very important debate around this has risen during the pandemic. See also the discussion in Vinuesa et al. (2020b)

Despite the importance of decentralised, trust-building approaches to development policy and the management of health crises, it is still the case, however, 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.<sup>12</sup> 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.

Thus, 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 General Data Protection Regulation<sup>13</sup> – 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% 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 initiative to regulate data ownership and usage in Africa, has only (so far) been ratified by eight member states.<sup>14</sup>

## Concluding remarks

In recent years, data-driven decision making has become increasingly important to inform policies that aim to improve global development outcomes, including those promoting the Sustainable Development Goals and Climate Action. The growing importance

of data gathering and use for global development is reflected also in the fact that the World Bank’s 2021 *Global Development Report* is to be devoted to data, entitled (at the time of writing) ‘Data for Better Lives.’

Concurring that how data is gathered and used may matter for ‘better lives,’ we argued that there are at least seven lessons for data-driven decision making for global development. What these seven lessons taken as a whole reflect, is that *how* human society chooses to govern data and digital technologies matters. There are three dimensions to such governance that given the current concerns are worthwhile stressing, namely cooperation, data-ownership and preparedness.

First, for the seven lessons drawn out here to enable better governance of data and digital technologies will require far better global cooperation and coordination – as well as cooperation and coordination between various regional governments – than what we have so far seen during the COVID-19 pandemic. Similar sub-optimal coordination scars global development in many areas, from fighting climate change, to global coordination on trade, investment and intellectual property, to name but a few. During COVID-19, the lack of global cooperation and sensible (non-herding) coordination in responses, including in the gathering and use of data, divisions over the role of the World Health Organisation (WHO)<sup>15</sup> and, moreover, in the development and roll-out of vaccines, have raised concerns 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).

Second, the seven lessons discussed in this paper would be much better taken on board if progress can be made with the global establishment and protection of data-ownership. This is not only to protect citizens for privacy-invading practices such as data harvesting by corporations, but to avoid mission creep from the extensive surveillance powers that many governments claimed during the pandemic. The danger is that without respect for data-ownership and privacy that once the pandemic is over, governments would continue to keep intrusive tabs on their populations (Harari, 2020). They can even potentially use the data obtained in the fight against COVID-19 for other, nefarious, purposes.

Third, there would perhaps have been fewer lessons learned the hard way during the COVID-19 pandemic if governments were better prepared. The COVID-19 pandemic was not a ‘black swan’ (Avishai, 2020). There were plenty of deep historical precedents, as well as a growing understanding of the likelihood of a global pandemic. Many Asian countries managed much better as a result of being better prepared after the earlier SARS, H5N1, H1N1 epidemics. 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.’

The seven lessons we discussed in this paper may be useful to improve global development policy and help in the management of a future global health crisis. However, these lessons are far from exhaustive, nor does experience or history negate the need for prudent global development policy to prioritise risk management. This means that the governance of data for global development should be aligned with the objective 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,<sup>16</sup> the COVID-19 pandemic has exposed global society as a ‘complex weak-link society.’ Data-driven decision making will only be an effective pillar of progress and resilience in global development if it can overcome and limit the structural weaknesses highlighted in this paper, including human cognitive biases, data gaps and disparities, digital divides and the related inequalities that these are often associated with.

### Acknowledgements

We are grateful to the editors and three anonymous referees for their comments and suggestions. Thanks also to Kunal Sen, Lorraine Telfer-Taivainen and Timothy Shipp from UNU-WIDER, who commented on an earlier version. Ricardo Vinuesa acknowledges support from the Swedish Research Council (VR). All errors and omissions remain our own responsibility.

### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Wim Naudé  <https://orcid.org/0000-0003-2496-6945>

Ricardo Vinuesa  <https://orcid.org/0000-0001-6570-5499>

### Notes

1. 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.
2. See <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>
3. Our description or definition of an infodemic is broader to that of Zarocostas (2020) who limits his discussion to misinformation.
4. Data science competitions have quickly grasped on the data deluge to issue challenges to data scientists to come up with solutions to manage the pandemic, see for instance Kaggle’s COVID-19 Open Research Dataset Challenge (CORD-19), at <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>.
5. 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 data used.
6. See <https://factcheck.afp.com/hot-air-saunas-hair-dryers-wont-prevent-or-treat-covid-19>.
7. Leon et al. (2020: e81) recommend that the best measure to use is weekly excess deaths, which would be ‘of deaths by all causes combined, thus side-stepping issues of what is or is not a death attributable to COVID-19.’
8. 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).
9. This happened in the past as well. In the fourteenth century, the Black Death boosted the market dominance and wealth of a few well-positioned incumbents (Russell and Parker, 2020).
10. The EU and the European Data Protection Board consider the decentralised collection and storage of personal data to more consistent with the General Data Protection Regulation (GDPR)’s principle of data minimisation, which requires the processing of personal data to be ‘adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed’ – See GDPR, Article 5.1c at <https://gdpr-info.eu/art-5-gdpr> than the centralisation of personal data.
11. At the time of writing, at least four privacy-preserving decentralised methods or protocols for contact tracing have been proposed, namely DP-3T, TCN Coalition,

- PACT (MIT) and PACT (UW) methods. The DP-3T (Decentralised 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.
12. See [https://www.freedomofthenet.org/sites/default/files/2019-11/11042019\\_Report\\_FH\\_FOTN\\_2019\\_final\\_Public\\_Download.pdf](https://www.freedomofthenet.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf)
  13. See the European Union General Data Protection Regulation (2016). Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
  14. See <https://au.int/en/treaties>.
  15. According to Fidler (2020), '...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 politicised COVID-19 for geopolitical purposes bodes ill for international health cooperation.'
  16. See <https://brandondonnely.com/2020/04/12/malcolm-gladwell-on-the-world-after-covid19>

## References

- Acemoglu D, Makhdoumi A, Malekian A, et al. (2019) *Too Much Data: Prices and Inefficiencies in Data Markets*. NBER Working Paper 26296. Cambridge, MA: NBER.
- Adam D (2020) Modelling the impact: The simulations driving the world's response to COVID-19. *Nature* 580.
- Aksoy CG, Eichengreen B and Saka O (2020) *The Political Scar of Epidemics*. NBER Working Paper 2740. Cambridge, MA: NBER.
- Alamo T, Reina D, Mammarella M, et al. (2020) Covid-19: open-data resources for monitoring, modeling, and forecasting the epidemic. *Electronics* 9: 827.
- Algan Y and Cahuc P (2010) Inherited trust and growth. *American Economic Review* 100: 2060–2092.
- Asencio-Cortés G, Morales-Esteban A, Shang X, et al. (2018) Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure. *Computers & Geosciences* 115: 198–210.
- Aubrecht P, Essink J, Kovac M, et al. (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.
- Avery C, Bossert W, Clark A, et al. (2020) *Policy Implications of Models of the Spread of Coronavirus: Perspectives and Opportunities for Economists*. NBER Working Paper 27007. Cambridge, MA: NBER.
- Avishai B (2020) The pandemic isn't a black swan but a portent of a more fragile global system. *The New Yorker*, 21 April. Available at: [www.newyorker.com/news/daily-comment/the-pandemic-isnt-a-black-swan-but-a-portent-of-a-more-fragile-global-system](http://www.newyorker.com/news/daily-comment/the-pandemic-isnt-a-black-swan-but-a-portent-of-a-more-fragile-global-system) (accessed 8 April 2021).
- Awad E, Dsouza S, Kim R, et al. (2018) The moral machine experiment. *Nature* 563: 59–64.
- Backhaus A (2020) 'Common pitfalls in the interpretation of COVID-19 data and statistics. *Intereconomics* 55: 162–166.
- Ball P and Maxmen A (2020) 'The epic battle against coronavirus misinformation and conspiracy theories. *Nature* 581: 371–374.
- Barocas S and Selbst A (2016) Big data's disparate impact. *California Law Review* 104: 671–732.
- Bergemann D and Bonatti A (2019) *The Economics of Social Data: An Introduction*. Cowles Foundation Discussion Paper 2171R. New Haven, CT: Yale University.
- 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 lock-down appear to be wrong, the same of the policies they generated. *Health Services Research and Managerial Epidemiology*. <https://doi.org/10.1177/2333392820932324>
- Brennen JF, Simon P, Howard, et al. (2020) *Types, Sources, and Claims of COVID-19 Misinformation*. Oxford: University of Oxford. Available at: <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation> (accessed 28 August 2020).
- Brynjolfsson E, Horton J, Ozimek A, et al. (2020) *COVID-19 and Remote Work: An Early Look at US Data*. NBER Working Paper 27344. Cambridge, MA: NBER.
- Burke M and Lobell D (2017) Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences of the United States of America* 114(9): 2189–2194.
- Bursztyjn L, Rao A, Roth CP, et al. (2020) *Misinformation During a Pandemic*. NBER Working Paper 27417. Cambridge, MA: NBER.
- Callaghan S (2020) COVID-19 is a data science issue. *Patterns* 1: 1–3.
- Campbell D and McIntyre N (2020) 'NHS data reveals 'huge variation' in Covid-19 death rates across England'. *The Guardian*, 13 July. Available at: [www.theguardian.com/world/2020/jul/13/nhs-data-reveals-huge-variation-in-covid-19-death-rates-across-england](http://www.theguardian.com/world/2020/jul/13/nhs-data-reveals-huge-variation-in-covid-19-death-rates-across-england) (accessed 8 April 2021).
- Casado M, Glennon B, Lane J, et al. (2020) *The Effect of Fiscal Stimulus: Evidence from COVID-19*. NBER Working Paper 27576. Cambridge, MA: NBER.
- Chavarria-Miró G, Anfruns-Estrada E, Guix S, et al. (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, Hua X and Maskus K (2020) *International Protection of Consumer Data*. CESifo Working Paper 8391. Munich: CESifo.
- Cirillo P and Taleb N (2020) Tail risk of contagious diseases. *Nature Physics* 16: 606–613.
- Desmet K and Wacziarg R (2020) *Understanding Spatial Variation in COVID-19 across the United States*. NBER Working Paper 27329. Cambridge, MA: NBER.
- Devarajan S (2013) Africa's statistical tragedy. *Review of Income and Wealth* 59: S9–S15.
- Ducharme L, Tebrake J and Zhan Z (2020) Keeping economic data flowing during COVID-19. *IMF Blogs*, 26 May.
- Duflo E (2017) The economist as plumber. *American Economic Review* 107(5): 1–26.

- Egger E-M, Jones S, Justino P, et al. (2020) Africa's lockdown dilemma: High poverty and low trust. UNU-WIDER Working Paper 2020/76. Helsinki: UNU-WIDER.
- Erondu N and Hustedt J (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](http://www.thenewhumanitarian.org/opinion/2020/06/18/COVID-19-policy-data-economy-health) (accessed 24 August 2020).
- Farboodi M and Veldkamp L (2020) Long-run growth of financial data technology. *American Economic Review* 110(8): 2485–2523.
- Feldstein S (2019) *The Global Expansion of AI Surveillance*. Working Paper. Washington, DC: Carnegie Endowment for International Peace.
- Ferguson N, Laydon D, Nedjati-Gilani G, 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, Wymant C, Kendall M, et al. (2020) Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* (6491): 368.
- Fidler D (2020) The World Health Organization and pandemic politics. Think Global Health, 10 April. Available at: [www.thinkglobalhealth.org/article/world-health-organization-and-pandemic-politics](http://www.thinkglobalhealth.org/article/world-health-organization-and-pandemic-politics) (accessed 24 August 2020).
- Furceri D, Loungani P, Ostry JD, et al. (2020) *COVID-19 Will Raise Inequality if Past Pandemics Are a Guide*. *VOX CEPR Policy Portal*. London: CEPR.
- Gad M, Elshehaly M, Gracanic D, et al. (2018) A tracking analyst for large 3D spatiotemporal data from multiple sources (case study: tracking volcanic eruptions in the atmosphere). *Computers & Geosciences* 111: 283–293.
- Giest S and Samuels A (2020) “For good measure”: Data gaps in a big data world. *Policy Sciences* 53: 559–569.
- Greenstone M and Nigam V (2020) Does social distancing matter? Becker Friedman Institute for Economics Working Paper 2020-26, University of Chicago, USA.
- Gruenwald E, Antons D and Salge T (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, Lau E, Wu P, et al. (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353: 790–794.
- Hilbert M (2016) ‘Big data for development: A review of promises and challenges. *Development Policy Review* 34(1): 135–174.
- Holpuch A (2020) US's digital divide “Is Going to Kill People” as Covid-19 exposes inequalities. *Guardian*, 13 April.
- Iacobucci G (2020) Covid-19: Deprived areas have the highest death rates in England and Wales. *British Medical Journal* 369.
- Ingelsby T and Haas B (2017) Ready for a global pandemic? The Trump administration may be woefully underprepared. *Foreign Affairs*, 21 November.
- Iverson T and Barbier E (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.
- 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).
- Knittel C and Ozaltun B (2020) *What Does and Does Not Correlate with COVID-19 Death Rates*. NBER Working Paper 27391. Cambridge, MA: NBER.
- Kremer M (1993) The O-ring theory of economic development. *The Quarterly Journal of Economics* (108): 551–575.
- Kuebart A and Stabler M (2020) Infectious diseases as socio-spatial processes: The COVID-19 outbreak in Germany. *Journal of Economic and Social Geography* 12429.
- Lazer D, Kennedy R, King G, et al. (2014) ‘The parable of Google flu: Traps in big data analysis. *Science* 343(6176): 1203–1205.
- Leszczynski A and Zook M (2020) Viral data. *Big Data & Society* 7(2): 205395172097100.
- Leon D, Shkolnikov V, Smeeth L, et al. (2020) COVID-19: A need for real-time monitoring of weekly excess deaths. *The Lancet* 395: e81.
- Luengo-Oroz M, Hoffmann P, Bullock J, et al. (2020) Artificial intelligence cooperation to support the global response to COVID-19. *Nature Machine Intelligence* 2: 295–297.
- Mehra M, Desa S, Ruschitzka F, et al. (2020) Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)31180-6](https://doi.org/10.1016/S0140-6736(20)31180-6)
- Mozur P, Zhong R and Krolik A (2020) In coronavirus fight, china gives citizens a color code, with red flags. *New York Times*, 1 March.
- Naudé W (2020) Artificial intelligence vs COVID-19: Limitations, constraints and pitfalls. *AI & Society* 35(3): 761–765.
- Naudé W and Cameron M (2021) Failing to pull together: South Africa's troubled response to COVID-19. *Transforming Government: People, Process and Policy*. <https://doi.org/10.1108/TG-09-2020-0276>
- NHS (2020) *COVID-19*. London: National Health Service. Available at: <https://covid19.nhs.uk> (accessed 24 August 2020).
- Ortutay B and Klepper D (2020) ‘Virus outbreak means (mis) information overload: How to cope. *AP News*, 22 March.
- O’Sullivan D, Gahegan M, Exeter DJ, et al. (2020) ‘Spatially explicit models for exploring COVID-19 lockdown strategies. *Transactions in GIS* 24(4): 967–1000.
- Paul S and Chowdhury P (2020) ‘Strategies for managing the impacts of disruptions during COVID-19: An example of toilet paper. *Global Journal of Flexible Systems Management* 21: 283–293.

- Pelizza A (2020) “No disease for the others”: How COVID-19 data can enact new and old alterities. *Big Data & Society* 7(2).
- Pestre G, Letouzé E and Zagheni E (2020) ‘The ABCDE of big data: Assessing biases in Call- Detail records for development estimates. *The World Bank Economic Review* 34(S1): S89–S97.
- Ramsetty A and Adams C (2020) Impact of the digital divide in the age of COVID-19. *Journal of the American Medical Informatics Association* 27(7): 1147–1148.
- Restrepo-Estrada C, Andrade S, Abe N, et al. (2018) Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computers & Geosciences* 111: 148–158.
- Rossello S and Dewitte P (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 Parker M (2020) How pandemics past and present fuel the rise of mega-corporations’. *The Conversation*, 3 June.
- Sandvik KB (2020) “Smittestopp”: If you want your freedom back. *Big Data & Society* 7(2): 205395172093998.
- Shen C, Taleb N and Bar-Yam Y (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).
- Serajuddin U, Uematsu H, Wieser C, et al. (2015) *Data Deprivation: Another Deprivation to End*. Policy Research Working Paper 7252. Washington, DC: World Bank Group.
- Taylor L (2020) The price of certainty: How the politics of pandemic data demand an ethics of ‘care’. *Big Data & Society* 7(2): 205395172094253.
- Tsikala Vafea M, Atalla E, Georgakas K, et al. (2020) Emerging technologies for use in the study, diagnosis, and treatment of patients with COVID. *Cellular and Molecular Bioengineering* 19.
- Van Dorp L, Acman M, Richard D, et al. (2020) Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infections, Genetics and Evolution* 83: 104351.
- Velásquez N, Leahy R, Johnson Restrepo N, et al. (2020) Hate multiverse spreads malicious COVID-19 content online beyond individual platform control. *ArXiv*, (2004) 00673.
- Vinuesa R, Azizpour H, Leite I, et al. (2020a) The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications* 11(233).
- Vinuesa R, Theodorou A, Battaglini M, et al. (2020b) A socio-technical framework for digital contact tracing. *Results in Engineering*. DOI: 10.1016/j.rineng.2020.100163
- Watts G (2020) COVID-19 and the digital divide in the UK. *The Lancet* 2(8): E395–E396.
- 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, Andrews J and Heaton T (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.
- Zarocostas J (2020) How to fight an infodemic. *The Lancet* 395(10225): 676.