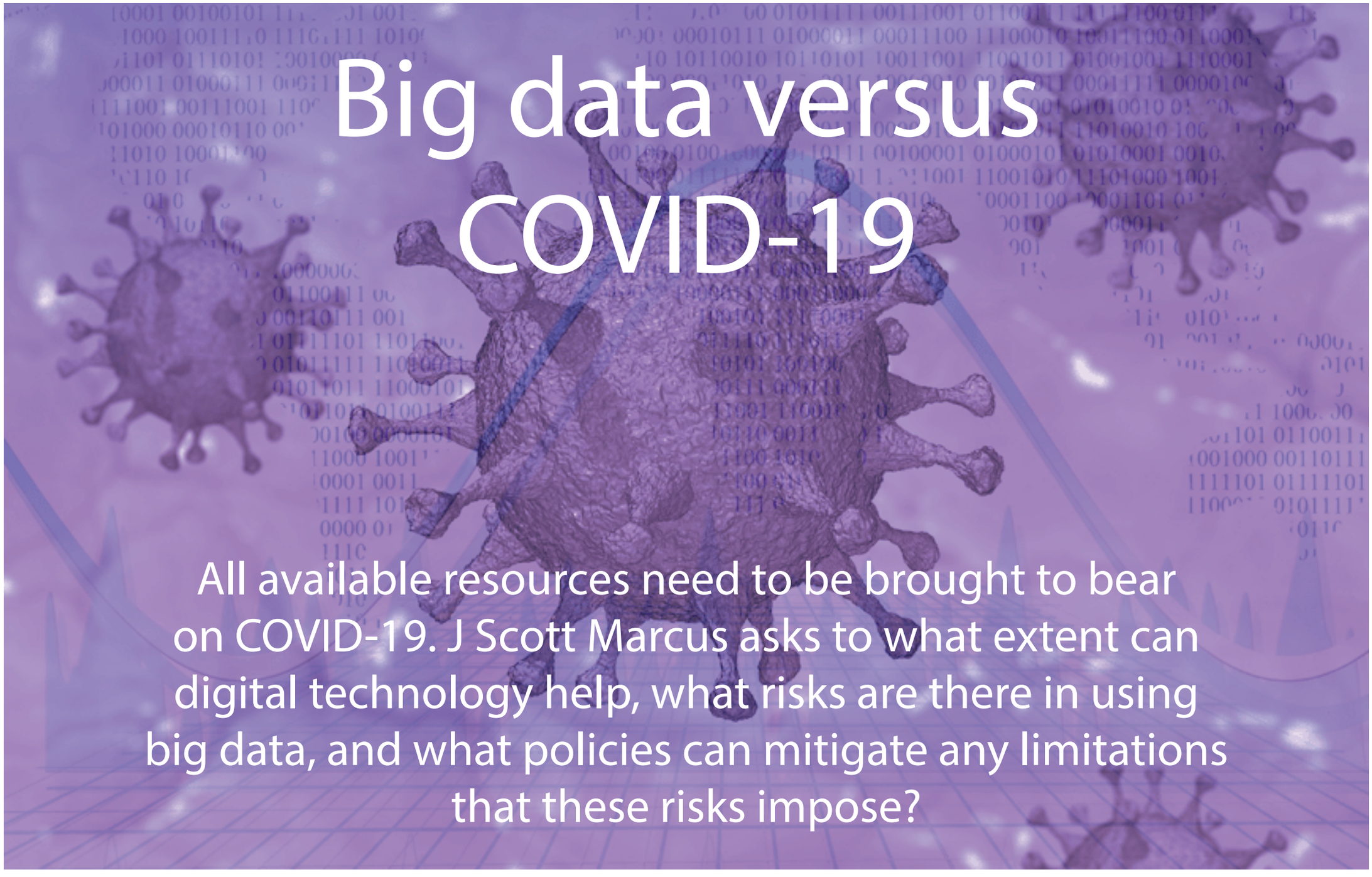
www.worldcommercereview.com

Big data versus COVID-19



All available resources need to be brought to bear on COVID-19. J Scott Marcus asks to what extent can digital technology help, what risks are there in using big data, and what policies can mitigate any limitations that these risks impose?

Information that could be made available

A great deal of COVID-19-relevant information is potentially available in the digital world:

- Users of social networks voluntarily provide extensive personal information, usually including demographics (age, sex) and location;
- Users of mobile networks provide information necessary to receiving and paying for the service, and also provide location information;
- Consumers who seek health information might voluntarily provide additional information.

Location data from mobile devices has been an area of intense interest for governments in the past few weeks. The mobile network knows your location, whether you are in your home country or roaming internationally.

Many countries have worked with the providers of communication services and infrastructure to progressively improve this location information, primarily as a means of improving the accuracy with which mobile users can call for help in the event of emergencies (see Marcus, 2010; Marcus, 2014).

Privacy challenges

However, use of personally identifiable data is restricted in most democratic, developed countries. The European Union implements the General Data Protection Regulation (GDPR) (European Union, 2016), which is based on the recognition of individual privacy as a human right.

That the EU has adopted a coherent overall horizontal framework for privacy is generally positive; however, the framework is relatively inflexible. This lack of flexibility becomes obvious now, when a nimble response is needed to a deep threat to the lives and safety of Europeans.

The use of data that is not personally identifiable is in general unrestricted, and several legal instruments at EU level actively encourage the making available of non-personal data and public sector information as a means of promoting economic efficiency (European Union, 2018 and 2019).

For commercial use of personally identifiable data, the GDPR puts a number of common-sense rules in place. The user must be told how the data will be used, to which third parties it will be provided and how they will use the data, how long data will be retained, and more.

Big data can play a constructive role in many different ways in helping the EU address the COVID-19 crisis

The GDPR's scope does not cover use of personally identifiable data collected by governments for purposes of law enforcement, which is a member-state competence.

Common practice in most developed democratic countries involves some combination of these elements:

- Data that is not personally identifiable (including anonymised data), or non-personal data, is subject to few if any restrictions.

- In order to collect data that is personally identifiable but that contains no content, public authorities must meet a fairly modest standard of proof of need. This tends to be the case for call data records (an indication as to who has been called from a telephone or internet device) and for user location data.

- In order to collect data that is personally identifiable and that contains actual content, a fairly high standard of proof of need must be met. Typically, an independent third party such as a magistrate must be convinced that there are valid grounds to suspect the individual, for instance of a past or likely future crime.

In order to understand how these broad principles interact with likely needs in terms of combatting COVID-19, it is useful to reflect on some of the use cases in which big data has been applied.

Ways in which big data has been used to date

There are three main forms of use that have been prominent to date: (1) strategic planning; (2) the tracking of (possibly infected) individuals; and (3) the provision of advice to concerned and possibly infected individuals.

Strategic planning

One of the most immediate and most promising uses of big data in combatting COVID-19 has been as a means of prediction, analysis, and strategic planning for national governments and national health authorities.

Epi-risk, for example, is a predictive model that looks at how the disease moves from one city to another as a function of air travel and commuting patterns. It draws on statistics on the number of known cases and deaths provided by national authorities, and on integration with air traffic data provided by the OAG database. The hope is that additional data from social networks can also be integrated.

According to the lead researcher, *"What we do as computer scientists and computational epidemiologists is provide [the doctors, nurses, and public health people in the field] with intelligence to anticipate the move of the enemy"* (Waltz, 2020a).

As another example, an analysis of the evolution of the disease in China (Li *et al* 2020) may serve to clarify the degree to which individuals who were not known to be infected contributed to the spread of the disease. The authors found that undocumented cases (ie. cases that had not been reported) were only half as contagious as documented cases.

Nonetheless, because some 86% of cases probably went unreported, they estimated that between 82% and 90% of all Chinese cases nationwide from 10–23 January were infected by people whose infections were undocumented. To estimate mobility between Chinese cities around Chinese New Year (which was 25 January in 2020), the researchers extrapolated from 1.7 billion records of 2018 travel records recorded by e-commerce merchant TenCent. This serves to demonstrate that big data can play a crucial role in valuable analyses.

Strategic planning in Austria, Italy and Germany has used mobile location data provided by mobile network operators. Mobility data from Deutsche Telekom is used to estimate the degree to which the German population is complying with requests or orders to stay at home.

In Italy, data provided by mobile network operators Telecom Italia, Vodafone and WindTre demonstrates that movements exceeding 300-500 metres in the Lombardy region are down by some 60% since 21 February, the date on which the first case in the region was identified. In Austria, A1 Telekom Austria Group is feeding mobility data into a third-party tool that is more typically used to estimate how crowded a ski area will become, but in this case can be used to estimate the effectiveness of social distancing (Reuters, 2020).

A common feature of all these strategic uses of big data is that they generally do not rely on personally identifiable data, or use anonymised data. This avoids most if not all privacy concerns.

This approach can be said to be much *"less invasive than the approach taken by countries like China, Taiwan and South Korea, which use smartphone location readings to trace the contacts of individuals who have tested positive or to enforce quarantine orders."* (Reuters, 2020).

Indeed Austrian privacy advocate Max Schrems observed that, *"As long as the [mobile location data] data is properly anonymized, this is clearly legal."* (Reuters, 2020).

Tracking of individuals

Taiwan is generally credited as having implemented effective measures to contain COVID-19, despite being among the countries initially thought to be most at risk. They had learned valuable lessons from the severe acute respiratory syndrome (SARS) epidemic in 2003, and benefitted from a high level of preparedness.

One of the most effective sets of measures was a system to track individuals thought to possibly be infected. As early as 27 January, the responsible government agencies integrated data about the past 14-day travel histories of individuals thought be at risk because of their travel history, with information linked to their health identification cards.

Individuals at high risk because of recent travel to affected areas were monitored electronically through their mobile phones. All hospitals, clinics, and pharmacies in Taiwan were given access to the travel histories of potential and actual patients (Wang *et al* 2020; Waltz, 2020b).

If measures like these were attempted in Europe, they might raise far greater concerns than the strategic measures because the individuals must be individually identified, and because the measures involve combining data normally collected for different and unrelated purposes.

This has been a central concern for an experimental system implemented by Oxford University. The plan is for individuals to download an app that would provide their location to the UK National Health System. The project would thus be voluntary, unlike in China, and the UK government has said it will delete the date and will *"not make the movements of infected individuals fully public, as has been done in South Korea"* (Valentino-DeVries, 2020).

Advice to possibly infected individuals

In contrast to these relatively successful stories, the bungled announcement of a website to *"sharply expand testing for the virus"* by President Trump on 13 March clearly showed the need to properly manage expectations and to address possible problems in perception. In a press conference, Trump incorrectly said that Google would provide a website to enable diagnosis of COVID-19 at scale.

Trump seemed to be referring to a small pilot project for the San Francisco area being worked on by Verily, a life sciences subsidiary of Google's parent company Alphabet. The work is at a very early stage. The Trump announcement was so far off the mark that Google felt obliged to issue a prompt correction/retraction (Shear and Wakabayashi, 2020).

The announcement needs to be understood as a political response to criticism of a US administration that has been under intense fire for allegedly having failed for too long to take the COVID-19 crisis seriously.

Verily was immediately deluged with requests, which generated a wave of criticism. A key concern, one of many, was that users had to log on using a Google account in order to access the site at all. This immediately raised concerns about possible use of personal data for advertising and for other unwanted commercial purposes (Wakabayashi and Singer, 2020).

Legal considerations

The use of big data in fighting COVID-19 is unlikely to be held up by legal obstacles:

- Strategic use of big data places little or no reliance on personally identifiable data. It is in consequence unproblematic.

- Under the EU treaties, health is a shared competence where the EU supports or complements the member states. If there is a tension between EU public health and privacy rules, one might reasonably expect privacy rules to take a back seat for a limited period of time.

- Many EU countries are already operating under conditions of national emergency, permitting governments to bypass normal legal protections for the duration of the national emergency.

Indeed, the European Data Protection Board (EDPB) made a public statement on 19 March 2020 that is fully in line with these principles (EDPB, 2020). They note that the GDPR already permits competent public health authorities and employers to process personal data when necessary for reasons of substantial public interest in the area of public health, as is the case during an epidemic.

Emergency measures are permissible, but only for the duration of the emergency. Data subjects need to be informed about the main features of the processing activities that are being carried out. Adequate security measures and confidentiality policies need to be in place. Anonymous use of mobile location data is permissible; however, the use of personally identifiable mobile location data should be avoided if possible, and if used must be subject to appropriate safeguards.

The real impediments to the use of data to combat COVID-19 probably have little to do with the legality of the measures put in place; they have instead far more to do with the risk of a loss of public confidence if use of personally identifiable data over-reaches, if personally identifiable data is allowed to be re-purposed for other unrelated purposes, or if the rationale for the use of the data is poorly communicated.

Implications for public policy

The implications for EU public policy are fairly clear. Big data can play a constructive role in many different ways in helping the EU address the COVID-19 crisis.

The use of **non-personally identifiable data** (including aggregated and/or anonymous data) will tend to be unproblematic, but the rationale should nonetheless be clearly thought through and clearly communicated.

To the extent that **personally identifiable data** is employed (without prior informed consent of the individual), national public authorities should take great care to ensure that a number of key common sense conditions (which are fully in line with the GDPR) are fulfilled:

- Data may be used only for the justified purposes intended. Use of personally identifiable data for commercial purposes that would not otherwise be permitted without informed consent should be strictly prohibited. The use of data for unrelated public purposes (tax enforcement, for example) should likewise be eschewed.

- Any personally identifiable data should be carefully protected against intrusion by hackers using good cybersecurity technology.

- Needlessly broad data collection should be avoided, since it creates risks (eg. of identity theft).

- Retention periods should be carefully considered. During the current epidemic, it will be important to understand the risk of re-infection. Some data might also be useful in fighting future epidemics, so a short deletion period is not necessarily the most appropriate policy approach for all data of this type.

For personally identifiable data, it is especially important that there be a clear and accurate public statement from government about what data is being collected, why it is being collected, with whom (if anyone) it will be shared, how it will be secured, and how long it will be retained. Otherwise, there is a significant risk of loss of public confidence in the measures undertaken.

If EU researchers are afraid to collect necessary data for legitimate purposes, member states should be prepared to enable a sensible and flexible response. For example, a member state might promptly issue 'comfort letters' to reassure researchers they will not be prosecuted for good-faith use of personally identifiable data in the context of projects that constitute valid and valuable research.

At the EU level, the public statement just issued by the EDPB (EDPB, 2020) possibly provides all of the clarity that is needed at the moment. ■

J Scott Marcus is a Senior Fellow at Bruegel

References
European Data Protection Board (EDPB) (2020), "Statement on the processing of personal data in the context of the COVID-19 outbreak", 19 March 2020.
European Union (2016), On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, (Regulation (EU) 2016/679).
European Union (2018), On the (Regulation (EU) 2018/1807)
European Union (2019), Directive on open data and the re-use of public sector information (Directive (EU) 2019/1024).
Kim, MJ; and Denyer, S (2020), "South Korea is doing 10,000 coronavirus tests a day. The U.S. is struggling for even a small fraction of that.", Washington Post
Li, R et al (2020), "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)", Science, 10.1126/science.abb3221.
Marcus, JS (2013), "The need for PPDR Broadband Spectrum", study on behalf of TCCA.

Marcus, JS; Burns, J; Jervis, V; Wähllen, R; Carter, K; Philbeck, I; and Vary, P (2010): "PPDR Spectrum Harmonisation in Germany, Europe and Globally."
Reuters (2020), "European Mobile Operators Share Data for Coronavirus Fight."
Shear, M; and Wakabayashi, D (2020), "Trump Oversold a Google Site to Fight Coronavirus", New York Times.
Valentino-DeVries, J (2020), "Translating a Surveillance Tool into a Virus Tracker for Democracies", New York Times.
Wakabayashi, D; and Singer, N (2020), "Coronavirus Testing Website Goes Live and Quickly Hits Capacity: The site from Google's sister company, Verily, was rolled out to two Northern California counties in hopes of guiding people to local virus testing", New York Times.
Waltz, Emily (2020a), "How Computer Scientists Are Trying to Predict the Coronavirus's Next Moves: Alessandro Vespignani describes the computational fight against the COVID-19 epidemic", IEEE Spectrum.
Waltz, Emily (2020b), "Big Data Helps Taiwan Fight Coronavirus: How Taiwan used big data, new technologies and heavy handed government to control the spread of the coronavirus", IEEE Spectrum.
Wang, CJ; Ng, CY; Brook, RH (2020), "Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing", JAMA online.

This article was originally published on Bruegel