Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
1. Introduction

The spread of the COVID-19 novel coronavirus and its rapid escalation into a pandemic in the early months of 2020 marks the first truly major, widespread global health emergency of the information age. Other disease outbreaks in the preceding decades had either been small or, in the case of 2009’s H1N1 swine flu epidemic, had lower infection and death rates than initially feared (Butler, 2010). Early analysis of COVID-19 using the CDC’s Pandemic Severity Assessment Framework (Reed et al., 2013) suggests that the ongoing pandemic is more comparable in severity to the 1918 Spanish Flu than to any more recent disease outbreak (Freitas, Napimoga, & Donalisio, 2020), and is by far the most serious such event to occur in the decades since the introduction and widespread adoption of information technology and networked consumer devices. As a consequence, the policy response to COVID-19 in many countries has effectively become a testing ground for the viability and efficacy of approaches which use information and communications technology (ICT) to enable or enhance various aspects of public health provisioning and targeting.

The role played by ICT and digital devices and platforms in the COVID-19 pandemic will be an important topic for study for many years to come, as there are few aspects of this pandemic and the public- and private-sector responses to it which have not been profoundly impacted by digital technology. Teleworking technology has permitted some parts of the economy – including much academic research – to continue functioning through otherwise extensive national- and regional-level shutdowns. Digital communication and social media platforms have played a role in supporting citizens’ mental health and sustaining their social and family relationships through extended periods of social distancing and isolation, but have also become a battleground for competing narratives around the pandemic – with guidance from health authorities and experts often struggling for prominence with conspiracy theories and false or outdated information. As citizens have sought to understand the progress of the pandemic, data scientists working at various organisations have led efforts to compile and verify figures for infection, testing and mortality, and to make them accessible to the public online with effective data visualisations (Bernard et al., 2020; Dong, Du, & Gardner, 2020). Each of the challenges, opportunities, strengths and weaknesses associated with these roles of ICT in the COVID-19 response will no doubt precipitate a significant scientific literature in the months and years to come. Perhaps the most significant role of ICT in the pandemic, however, and certainly one of the most...
controversial and contested even at this early stage, is the experimental adoption of digital contact tracing and exposure notification – using citizens’ personal digital devices such as smartphones to trace their physical movements and interactions with other citizens, thus allowing citizens themselves or medical authorities to be notified when they have come into contact with infected individuals.

2. Contact tracing and digital technology

Contact tracing itself is not new – it is a well-established part of the response to any contagious disease outbreak. In an April 2020 media interview, CDC director Robert Redfield described “very aggressive” contact tracing of infected individuals as an essential step in bringing COVID-19 under control (Simmons-Duflin et al., 2020) – but while he noted that technological solutions to improve contact tracing were being evaluated, his focus was on the more traditional and enormously labour-intensive form of contact tracing, which requires a large number of public health fieldworkers to contact family, friends, coworkers and other contacts of infected individuals, arrange for them to be tested or quarantined, and interview them to find out about their potential contacts in turn. For a pandemic on the scale of COVID-19, even a medium-sized country could require tens of thousands of full-time fieldworkers to run a comprehensive contact-tracing program in this way.1 As such, large amounts of attention and resources have been focused on finding ways to leverage digital technology to automate significant parts of this process – taking advantage of the fact that a majority of citizens in developed countries (and large numbers in developing nations) carry smartphones which integrate GPS chips capable of precise location tracking, Bluetooth radios which can sense the proximity between devices, and always-on connections to the Internet. Digital contact tracing seeks to use this functionality to turn citizens’ own smartphones into contact tracing devices. The hypothetical advantages of this approach are significant; in an ideal world situation, it would allow contact tracing to be extended to a country’s entire population rather than just a subset of infected individuals, would track their movements and social contacts with a very high degree of precision, and would be faster, more efficient and less labour-intensive and prone to human error than existing approaches.

While the fundamental objective of digital contact tracing applications is relatively straightforward, however, different countries have taken quite different approaches to the development, roll-out and functioning of such applications. Initially, this divergence was largely just a facet of the more general piecemeal approach to COVID-19 response among nations; by mid-March 2020, contact tracing applications of various kinds had been developed and rolled out independently by authorities in Israel, Singapore, South Korea, Taiwan, Thailand and Vietnam (Cho, Ippolito, & Yu, 2020; Lee, 2020), pre-empting the announcement on April 10, 2020 by smartphone vendors Google (Android) and Apple (iOS) that they were working jointly on a unified framework for contact tracing which would be built directly into the operating systems of phones powered by their software. Even after Google and Apple made their solution available, however, many countries continued to pursue the development of their own contact tracing applications, which would bypass the Google / Apple framework entirely. This has revealed that although the piecemeal nature of early efforts to develop contact tracing applications was partially down to a lack of any coherent, centralised solution, there is also a fundamental difference in the attitudes of different countries and corpora
tions to how these applications should function, and, most crucially of all, who should have access to the data they generate. This has placed what looks ostensibly like a straightforward and positive technological advancement to help limit the spread of a deadly disease at the nexus of a turbulent and bitterly fought argument over citizens’ rights over their own digital data, online privacy and surveillance in the information age. While only a few years ago this debate might have been a somewhat abstract one carried out largely in the rarefied air of academic or public policy discourse, a litany of major scandals over the use of citizens’ digital data in illicit ways by companies such as Cambridge Analytica (Isaak & Hanna, 2018), escalating concerns over digital surveillance and espionage, especially by authoritarian governments, and high-profile clashes between politicians and public officials in several jurisdictions and the operators of social networks such as Facebook and Twitter, have given these concerns central prominence in the public sphere. The deep divide between different philosophies over digital contact tracing, data sharing and user privacy are therefore anything but abstract; they are reflections of real-world issues which have the potential to profoundly impact the efficacy and success of the technology – efficacy and success which will, in this instance, ultimately be measured in lives saved or lost.

3. Data first or privacy first?

While there are many fault-lines among nations and corporations in their differing approaches to contact tracing, perhaps the most salient of them is the divide between “data-first” approaches, which prioritise the retention of tracking data and its availability to health authorities and researchers, and “privacy-first” approaches which emphasise citizens’ control over their own data and seek to provide an effective degree of contact tracing without exposing identifiable individuals’ movements and interactions to authorities. From a data management perspective, the former approach (data-first) generally involves assigning a stable identifier to each individual (or smartphone device) and transmitting some or all details of their movements and contact interactions to a central server, where they can be accessed and analysed. The latter approach (privacy-first), in contrast, uses dynamic identifiers for individuals which are changed regularly, and stores their contact interactions in a cryptographically secure manner on their local device, keeping little or no data in a centralised server.

The most basic level of functionality enabled by these two approaches is the same – an alert can be issued across the network when an individual tests positive for COVID-19, either by the health authorities directly issuing the alert or by the individual entering a specific code on their device. This alert will inform anyone who had a contact interaction with the infected person that they may have been exposed to the virus. Beyond this basic level functionality, however, there is an enormous divergence between the two approaches. The data-first approach potentially allows health authorities to directly identify and contact individuals who have come into contact with the virus, whereas the privacy-first approach does not identify individuals and only allows them to be notified on their smartphones, leaving the responsibility for contacting health authorities or submitting to a test up to the person themselves. The data-first approach also commonly includes GPS location data along with the contact interaction log, allowing health authorities to locate the specific venues where clusters of infections have occurred, something which is not possible with the privacy-first approach – though in theory, the privacy-first approach will still have notified the people potentially exposed to the infection cluster, albeit without telling them where the cluster occurred or from whom the infection may have been passed. Finally, the data-first approach generates a large quantity of data on the movement of and contacts between individuals and how it pertains to the spread of the virus through the population, making it into a potentially invaluable resource for epidemiologists and data scientists researching both the COVID-19 virus specifically and the mechanisms of epidemics more generally.

It is important to note that the data-first and privacy-first approaches to digital contact tracing described here are not binary opposites, but rather represent the extremes of a spectrum of different

---

1 For example, UK Health Secretary Matt Hancock announced on 23 April, 2020 that 18,000 people will be hired to trace the contacts of those infected (BBC News. 2020a, April 23).
approaches to the problem. Different approaches developed by different governments and corporations cover a wide range of the possible spectrum. South Korea, notably, has taken a dramatically “open” data-first approach; a combination of human contact tracing efforts and digital data, including GPS location, are used to create a “virus patient travel log” which is available in partially-redacted form to the public (Kim & Denyer, 2020). Singapore’s contact tracing system shares less personal information about infected individuals, but the government maintains a publicly accessible map with details of each case, raising the risk of individuals being rightly or wrongly identified as virus carriers (Raskar et al., 2020). The United Kingdom, France and Australia, among many other nations, prefer a data-first approach which avoids sharing the collected data with the public but still makes it available to health authorities – though even here there is divergence, with some countries keen to share the data with researchers and other interested parties, while others are enacting legal frameworks to forbid access to the data even in the case of court orders (BBC News, 2020b, April 26; Kelion, 2020). The frameworks developed by Google and Apple, which largely build on the idea conceived by the Decentralized Privacy-Preserving Proximity Tracing (DP-3 T) protocol and are being implemented in countries including Germany, Italy, Japan and many U.S. states, meanwhile, fall strongly into the privacy-first end of the spectrum, creating no accessible archive of contact or location data and entirely concealing users’ identities. The distribution of these approaches across different countries depends on a variety of complex factors; while it is tempting to seek a single explanatory variable such as the country’s degree of political freedom, it is notable that the most strongly data-first approaches have been adopted by liberal democracies such as South Korea and Taiwan as well as authoritarian states like China, Iran and Qatar, while both France and the UK have also resisted the privacy-first approach (Ó’Neill, Ryan-Mosley, & Johnson, 2020). Asia in particular offers a most divergent set of data-first and privacy-first countries; among other factors in this decision, we might also need to think of these nations’ previous experiences with SARS and MERS and the timing of democratisation.

The complexity created by these very different national approaches to digital contact tracing is deepened further by very short timespan in which the COVID-19 pandemic has emerged, meaning that facts on the ground have been very fluid and best practices have been subject to rapid change, where they have existed at all. Some major nations have been forced to change their approach entirely; Germany initially favoured a data-first approach but later shifted entirely towards a privacy-first approach, while both the United Kingdom and Australia encountered significant technological impediments and problems with their data-first contact tracing applications, with the UK even commencing development of an entirely separate privacy-first application as a backup plan for the failure of its data-first approach (Selby, 2020). Similar technical problems have arisen in a number of countries which opted for data-first approaches, as this has necessitated ignoring Google and Apple’s operating system level frameworks in favour of developing proprietary applications which work around smartphones’ privacy and security features. One major consequence is that these applications tend to have a large impact on device battery life, which could result in users opting to turn them off or remove them from the device. There are questions, too, about the security and reliability of these applications – which rely to some degree on operating system loopholes and may not robustly implement cryptographic security or user anonymisation techniques (Deep, 2020). Japan, for example, was forced to shift from initially developing its own apps in concert with three local firms to a new approach based on Google and Apple’s frameworks, again due to concerns over compatibility with the operating systems of most mobile devices. COVID-19 serves as yet another case which reminds us how tenacious path-dependency is where the hardware of big technology companies is involved.

4. The risks of data over-collection

Given the pressure brought to bear by Google and Apple’s adoption of a strongly privacy-focused approach, and the changes being made to the approaches of countries such as Germany, the UK and Australia after encountering a variety of difficulties with their initial data-first approach, it seems clear that the direction of travel for digital contact tracing is towards the privacy-first model. Though this has been welcomed by privacy and security advocates, it has caused disquiet with some researchers and health authorities, who fear that their inability to access location data and pinpoint the origin of infection clusters will significantly degrade the usefulness of contact tracing.

This view hinges closely to the conventional wisdom that has been widespread across many fields of science – but especially the social sciences and information science – in recent decades. With the rapid advances made since the 1990s in the field of data science and the now widespread availability of tools for storing, handling and analysing formerly unimaginable volumes of data, the mantra of many researchers and public officials has become that you can never have too much data. This is a mantra shared with the private sector, where much of the value of companies like Facebook, Google and Amazon actually lies in the extraordinary amount of data which they possess about their users and customers – an ocean of individual data points which may individually be worthless, but assessed by machine learning algorithms and explored for underlying patterns and correspondences, can reveal enormous amounts about the behaviour of both individual users, and communities and societies as a whole. Viewed with this understanding, the privacy-first approach to digital contact tracing can appear hugely wasteful, as it disposes of (or simply never collects) vast amounts of data that could be of enormous value to improving public health policy countermeasures to COVID-19 as well as advancing our knowledge of epidemiology and its related fields.

As frustrating as it may be for researchers and health authorities watching this potential treasure trove of data being locked away and destroyed, it is important to remember that the data in question is the personal data of citizens – and in recording all of their contact interactions (and in some cases, all of their movements) it represents arguably the most personal and intimate data a government has ever sought to gather about its own citizens. It is entirely unsurprising that this would immediately run afield of privacy concerns which reassurances regarding the trustworthiness of government agencies, the limitations on the distribution and use of the data, and technical measures such as the anonymisation of identifying data have done little to assuage. The Cambridge Analytica scandal and a growing understanding of how personal data has been leveraged in marketing and election campaigns have been milestones in public awareness of digital privacy, but the underlying worries are not new. As early as the late 1990s, concerns were voiced over the “dossier effect” whereby the collection of large numbers of seemingly innocuous data points could create a combined data set with a startling amount of personal information that is easily de-anonymised and attached to an individual citizen (Goldberg, Wagner, & Brewer, 1997). Since then an extensive literature on varieties of digital privacy and methods for its protection has developed, but the problem remains a live and increasingly politically fraught issue. Mainstream public discourses on digital privacy have shown an increasing degree of suspicion in recent years, from the relatively widespread belief that digital assistant devices such as Amazon’s Alexa can “spy” on their users’ conversations, to widespread negativity about the social and political impact of companies like Facebook which overtly collect large amounts of user data (Newton, 2020).

A major consequence of this for COVID-19 contact tracing is that convincing citizens to actually install and use these applications presents a significant challenge, given the greater than ever level of awareness of privacy and personal data concerns. While some countries such as China, India and Qatar have legally mandated the use of the applications, this is not possible in most liberal democracies – and even
where it is mandated, users who fear their location being tracked by unwanted parties or even leaked to the public may exercise non-compliance simply by leaving their smartphone at home. Where usage is not mandatory, citizens who distrust any link in the chain of public- and private-sector bodies involved with the application – technology companies, governments, health authorities, private sector outsourcing firms, and so on – may choose not to install the application, or to regularly disable it, in order to protect their own privacy. High-profile cases where contact tracing has been involved or implicated in breaches of citizen privacy have already become well known, such as the potential “outing” of a number of gay men in South Korea (a country with almost no legal rights or protections for LGBT people) after a cluster of COVID-19 cases occurred in an area of Seoul famous for gay bars and clubs, or the claim by a Minnesota law enforcement official that the state was using “contact tracing” to identify the connections of protestors arrested during May’s Black Lives Matter demonstrations (Mullin, 2020). These incidents and others like them will only serve to deepen the concerns of many citizens that cluster tracing risks revealing their private information in ways that could ultimately be harmful for them – thus forcing them to weigh the new risks created by the cluster tracing app against the COVID-19 risks it supposedly helps to obviate.

The first goal for these applications, far ahead of any concerns about retention of data for research purposes or the secondary objectives of health authorities, must be widespread adoption. The applications are more or less useless from a public health perspective unless they are installed and used by a critical mass of citizens – which in liberal democracies means convincing citizens to install them, and even in illiberal countries requires citizens to exercise full compliance rather than trying to avoid tracking. Given the overall climate of concern over privacy rights, the melange of different approaches and the strong voices calling for data-first systems which require citizens to utterly trust the government authorities actually risk making digital contact tracking largely ineffective. Early research into these systems has noted that each application raises a large number of questions – such as how exactly they function, what data they’re storing and where they’re storing it, to whom they’re transmitting information and what the users’ rights are regarding their own information in the app (Kishimoto & Kudo, 2020) – which are often poorly or vaguely answered both by descriptions in software itself and by the authorities responsible for running the system. This lack of clarity and transparency, too, will inevitably lead to citizen non-compliance in digital contact tracing efforts – given falling levels of institutional trust across much of the world, many people glimpsing the debate playing out over the balance between data and privacy will simply assume the worst of their own government’s efforts. COVID-19 thus serves as a difficult trial for trust in public institutions both nationally and internationally. The glimmer of hope in all of this, perhaps ironically, is that many citizens, at least in the U.S., do appear to have a relatively high degree of trust in Google and Apple (Newton, 2020) – suggesting that an emphasis on their role in developing contact tracing, and the priority they have placed on user privacy, might help to assuage some of the fears that have been created by data-gathering overreach, boost citizen participation in these systems, and ultimately save lives.

5. Implications for research

The experience of health authorities seeking to rapidly develop and implement digital contact tracing applications in countries around the world should serve as a sobering case study for researchers whose work relies on gathering large amounts of information about individuals – both in public health and across many other fields. The wide divergence of different approaches has revealed a complex set of overlapping national and regional factors related to priorities in research data collection, strength of privacy protection both in law and in common perception, depth of public trust in various institutions, and the ability of national authorities to resist or avoid the path-dependency imposed by the dominance across many fields of a small number of technology companies. The fact that no global consensus on best practices for digital contact tracing has emerged and that approaches which appear to be working in some countries have proved entirely unavailable in other countries – even in neighbouring countries, in some cases – highlights the enormous and still poorly-understood complexity of this mesh of interacting institutional, legal, cultural and social factors, and how it restrains and shapes the possibilities of gathering, storing and analysing large amounts of citizen data.

As a result, there will be an urgent need – once the COVID-19 pandemic has ended – for a serious and robust assessment of which contact tracing protocols were effective and which were not, taking into consideration not only the variation between those protocols created by national- and regional-level factors, but also seeking to understand the extent of citizen compliance each of them succeeded in achieving. Researchers working with large-scale data have understood for many years that the “age of innocence” in data collection – be it public opinion surveys, census-taking, social media data or any form of activity logging – is behind us, replaced by an era in which the subjects of data collection are aware of the collection and analysis process, often suspicious of the motives behind it, and in some cases willing to deliberately attempt to alter their responses or measured behaviours in order to confuse the analysis or influence its results. Examples of this behaviour such as people giving false responses to public opinion surveys in order to influence the results, or using software to mask or alter their online identity and activity, are for the most part mere irritants, though they present a major challenge to data-focused researchers. Public suspicion directed at COVID-19 contact tracing applications, however, could potentially have a cost measured in lives. If large parts of the population refuse to participate in tracking, or deliberately obfuscate their movements in other ways, it indicates an enormous failure of public engagement and policy around the tracing system – one which also implies an ongoing problem with public data collection efforts of many types.

As large-scale events are wont to do, COVID-19 has revealed the boundaries of our knowledge and understanding in many fields – information management and science included. In doing so, it has made clear a number of important and urgent frontiers for research work. The global nature of the pandemic demands that we piece together the patchwork quilt of different factors that have determined the implementation and success rates of various approaches around the globe. Moreover, the collision of this public health crisis with the looming but still under-studied crisis in public faith in data security and privacy has made it more important than ever that we understand public attitudes towards these issues, both at a national and individual level – a question whose answers are undoubtedly complex and rapidly changing, requiring the use of extensive surveys and other approaches to start to approach a set of useable answers. Finally, the questions and difficulties raised in the implementation of contact tracing around the globe have made clear that authorities and private actors alike often lack sufficient understanding of public concerns about their privacy and the use of their data, and fail to provide clear, up-front answers to citizens’ concerns when asking them to provide data. Too many assumptions are made; authoritarian states assume that they can enforce compliance, while many liberal democracies have fallen into the trap of assuming a high level of public trust in institutions will ensure compliance, or even simply believing that the public doesn’t really care about digital privacy issues at all. Research must be focused on finding a way to reassure citizens and maximise compliance without making such assumptions – not only in order to improve and protect the integrity of public data collection and processing in general, but also to ensure that the confusion and suspicion which has greeted contact tracing applications in many countries can be overcome in the event of any future epidemic, such that this valuable tool is not entirely removed from the arsenal of public health authorities.
