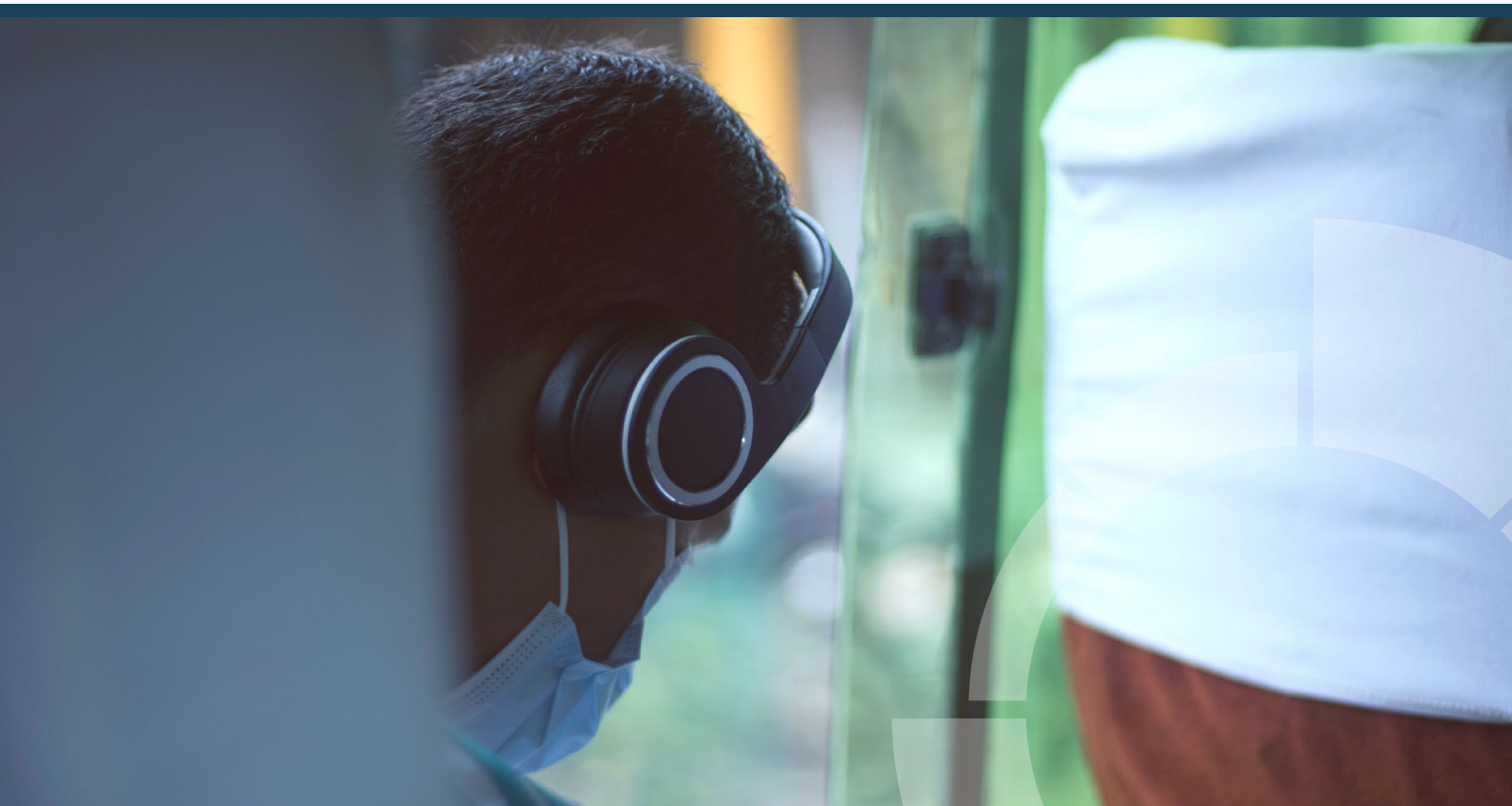
December 2020

THE TRINITY CHALLENGE

Better decisions to protect against health emergencies

We can and must make data-driven decisions to identify, respond to, and recover from outbreaks, epidemics and pandemics.



In brief

Globally, we are inefficient at getting the right people access to the right data at the right time. We are ineffective in combining insights gained across data sources and sectors to support the identification of, response to, and recovery from health emergencies. COVID-19 has demonstrated that our use of data to make decisions that improve health, economic and social outcomes is vastly, and unnecessarily, insufficient.

Everyone – private-, public- and social-sector leaders, as well as governments – must take immediate action to remedy this. This includes accelerating efforts to make better use of publicly available data, striving to increase the breadth and depth of access to data held by corporate firms (in a privacy preserving fashion), and bring data providers, analysts, researchers and decision makers into closer collaboration.

In the medium term, leaders can ignite a discussion around the need for data-governance and access protocols when responding to a health emergency, shift the paradigm from 'the value of data' to 'the value of outcomes', and shape the development of a playbook for decision makers.

It is imperative that we take stock of how we have generally failed to support efficient and effective decision making due to our lack of sophisticated and sensitive data analytics.

This is fundamentally a leadership challenge, not just a legal or regulatory one. Additionally, while advances in technology will help improve data-driven decision making, they are not sufficient. We can and must move forward together to improve data-driven decisions in the immediate response to this health crisis, and in anticipation of the next.

Contents

Introduction

The COVID-19 pandemic continues to wreak havoc on our lives and livelihoods. Many governments, institutions, businesses and communities across the world have taken bold and decisive action to protect lives and mitigate the economic impact of the pandemic – yet the events of the past year have revealed profound gaps in humanity's preparedness for health crises. It is clear that all actors can and should work together to do better next time.

Indeed, the COVID-19 crisis should provide impetus for discovery and improvement. Jawaharlal Nehru, India's first Prime Minister (and an alumnus of Trinity College) reflected that "crises and deadlocks when they occur have at least this advantage, that they force us to think."[1]

The potential rewards of better, more collaborative thinking and action are significant. What if, in future, the investments and infrastructure are in place to identify an emerging or re-emerging infectious-disease threat before it becomes an outbreak or epidemic? What if humanity learns to analyse rapidly which targeted, coordinated interventions are most effective in response to an infectious disease, and how to intervene at the lowest social and economic cost? What if researchers come to understand how to initiate a social and economic recovery in a way that does not exacerbate inequalities?

This discussion paper, itself a global collaboration between actors from academia, the health sector and business, aims to foster such joined-up thinking and action. It is published under the auspices of The Trinity Challenge, a coalition of members united by the common aim of better protecting the world against health emergencies, using data-driven research and analytics.

The authors point out that decision making in health emergencies – be it by government officials, business leaders, public-health agencies, civil society or medical bodies – relies increasingly on the sensitive and sophisticated analysis of data, and the appropriate application of the insights and information derived from that analysis. Of course, decisions to safeguard lives and livelihoods are not solely dependent on data; they are made in combination with political judgements and the resources that a particular system has available at any given moment. However, the authors are of the firm view that, in response to COVID-19, decisions have been made in conditions of unnecessary uncertainty. That is because data, and the insight data can provide to decision makers, have not been adequately utilised.

The initial part of this paper sets out the significant common challenges that decision makers across geographies, sectors and disciplines face in harnessing data to protect against health emergencies. These challenges are grouped into three categories: (i) *efficiency*, or ensuring that the right people access and analyse the right data at the right time; (ii) *effectiveness*, overcoming fragmentation and the creation of silos in the insights that arise; and (iii) *sufficiency*, ensuring that decision makers have the information they need.

The paper then turns to the immediate and concrete actions that institutions across sectors can take to begin to meet those challenges. These actions include accelerating efforts to make better use of publicly available data; increasing the breadth and depth of commercially held data that is efficiently accessible; and bringing data providers and decision makers into closer contact.

Finally, the paper looks to the longer term, arguing that leaders of institutions should act with the ideal system in mind, even in the urgent context of responding to a health emergency. That will entail advancing discussions on a more global system of data governance and access protocols; shifting the paradigm from 'the value of data' to 'the value of outcomes'; and shaping a robust playbook for decision makers.

The paper does not pretend to provide fully formed answers in these areas. Rather, its intention is to challenge the world's best and brightest minds to contribute ideas and innovations to better protect us against health emergencies, using data-driven research and analytics. In so doing, the paper – and the broader efforts of The Trinity Challenge – can help close significant gaps in the existing research base. It is also hoped that the principles outlined here can work to catalyse a shift in societies' approach to health emergencies, one that favours investigation over impulse, and solidarity over solitary action.

Inclusivity, innovation and collaboration are the principles that guide The Trinity Challenge as it stimulates inquiry and partnership across disciplines, sectors and geographies. In this paper, the authors advocate for a practical and powerful expression of these principles, one which harnesses current scientific and technological tools: the acquisition and analysis of data in highly collaborative networks. We show how a data-focused, collaborative approach would serve humanity as an essential means to identify, respond to, and recover from any emerging and re-emerging infectious-disease threat.

This paper is born from ideas raised in conversations with The Trinity Challenge's expert working group, individual interviews, and targeted outreach across a variety of industries. A snowball sampling methodology has been used to ensure representation from across geographies and relevant sectors. In proposing a data-driven approach to enable better decisions to improve health, economic and social outcomes, we hope to offer an understanding of a shared underlying problem, provide helpful common language and concepts, and encourage immediate actions and new findings.

Chapter 1.

Common data challenges across geographies, sectors and disciplines

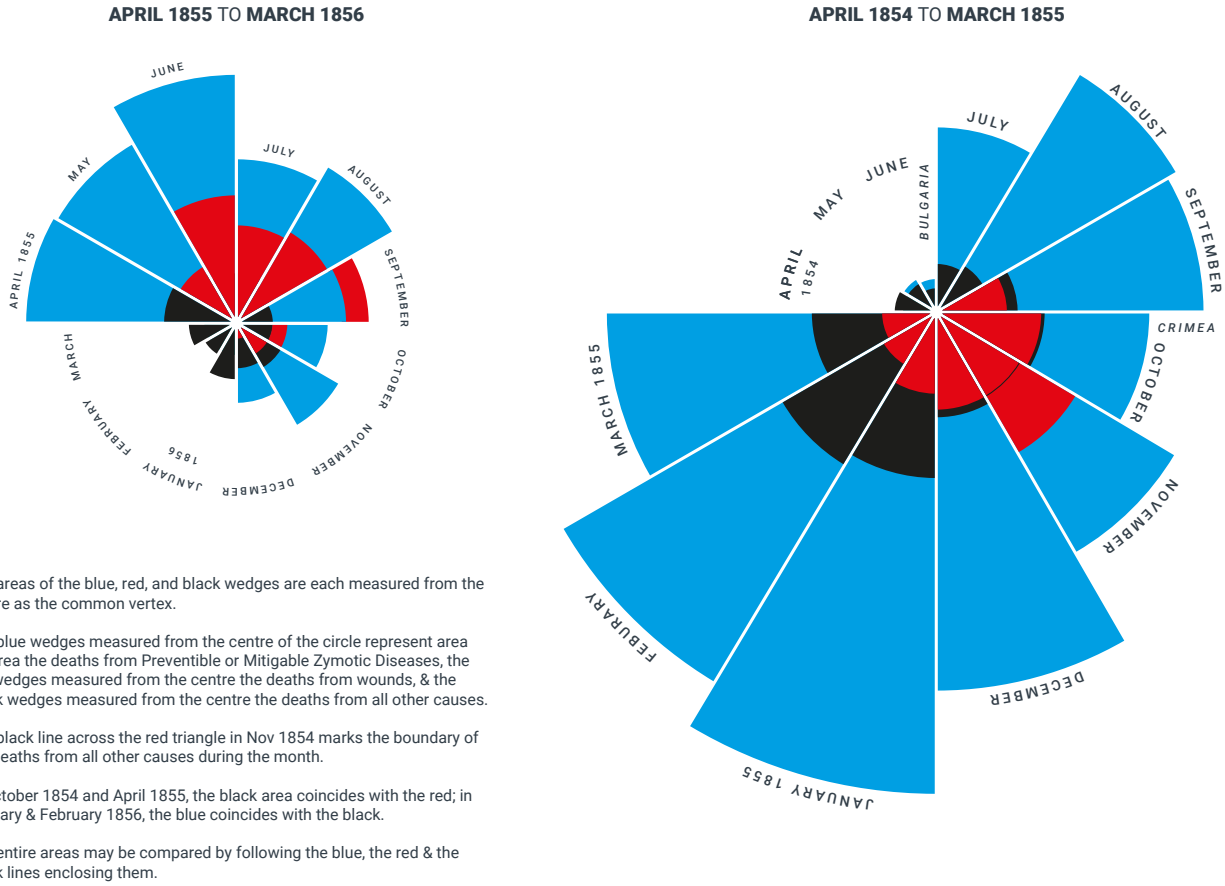
Decisions in response to the COVID-19 crisis have been made in conditions of unnecessary uncertainty, because relevant data, and the information that data provides, has not been optimally utilised. Those who develop and monitor the implementation of health and economic interventions – whether in the COVID-19 crisis or future pandemics – would benefit from access to a range of different data sources, including data relating to financial transactions and consumer spending, use of public transport, and the transmission of disease. Some of this data is publicly available, while some is held by private institutions where the data typically has commercial value.

As we shall show, we have significant opportunities not only to increase access to the right data, but also to ensure that the right people have access to the right data at the right time. Of course, methodologies and technologies will evolve over time, which means that analysis can also improve.

However, it is not only the analysis of data that matters, but the use of the resultant information to take better decisions.

One trailblazing public-health thinker who knew the importance of analysis and communication was Florence Nightingale. She revolutionised medicine not only with her advances in hygiene, sanitation and patient care, but also by means of data and analytics. Take her 'Rose Diagram' of 1858, a statistical graphic which showed that epidemic disease – which was responsible for more British deaths in the course of the Crimean War than battlefield wounds – could be controlled by factors including nutrition, ventilation and shelter (Figure 1). This was a simple, clear and persuasive explanation of complex statistics. Nightingale highlighted the value of translating research and evidence into data, analysing data into information, and using the information to communicate a decision and a clear message.

Diagram of the causes of mortality in the army in the east



The areas of the blue, red, and black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854 and April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Figure 1 – Florence Nightingale's Rose Diagram, on which The Trinity Challenge logo is based.

Data collectors and providers can be found across the globe: they operate at local, national and international levels; in public and private sectors, from government and healthcare to telecommunications and financial services; and in roles from data analysis to executive decision making. But COVID-19 has highlighted pervasive weaknesses in data-sharing systems. Among the lessons to be gleaned are three clear findings across the private, public and social sectors:

1. We – actors in all sectors – are inefficient in getting the right people access to the right data at the right time.
2. We are ineffective in combining insights gained across data sources and sectors.
3. This calls for more efficiency and effectiveness in our use of data and information to make fact-based decisions and improve outcomes.

Decision makers do not have the information they need to take decisions that improve outcomes

	Data →	Insights →	Decision-making
Challenges	**Inefficient** at getting the right people access to the right data at the right time	**Ineffective** information sharing as data analysis remains siloed and fragmented	**Insufficient** decisions as decision makers lack the critical information needed
What is needed	Representative and accurate data generation Unbiased data collection Privacy-preserving and secure access Real-time analysis	Shared information standards and protocols Timely and defragmented datasets, out of silos Aligned incentives to collaborate on data	Tighter feedback loops between data providers and decision makers Faster connections between analysts to signal required improvements in data collection Engagement and communication with public

Figure 2 – COVID-19 has highlighted the pre-existing weaknesses in our data-sharing systems.

COVID-19 is a powerful reminder of our inextricable shared humanity. Data and information can be better used to understand the links between us, and to protect and promote health, social cohesion and economies. Remedying the problems identified with analysis of data, use of information, and subsequent decisions will have benefits for many other elements of healthcare systems. In the end, everyone will benefit from data-driven decisions in health emergencies and beyond.

In regard to COVID-19 and future health emergencies, we shall offer two examples in this chapter to illustrate the benefits of doing better. The first involves the analysis of mobility data, enabling assessments of how the movement of people can contribute to disease spread in a population. The second relates to accessing and analysing monetary transactions, which would help us to estimate the effects of an outbreak or epidemic on economic outcomes.

Decision makers must understand whom the virus has infected, where that person is, and how and at what speed the virus is moving between people. COVID-19 incidence varies in different locations and is typically derived from official case numbers. This data is usually delayed, and tells decision makers very little about what is likely to happen in the future. High-quality origin-to-destination mobility data could be used to refine these estimates, enabling more targeted policy responses at a local level. Indeed, mobility data is critical to guiding public-health actions in the early, middle

and late phases of outbreaks, epidemics and pandemics (see **Case study 1**).[2] This data is available from mobile-phone companies, which know about movement because it corresponds to customer use of different cell-phone towers.

Decision makers must also understand economic impacts. Government statistics of household consumption and economic growth are usually used by policymakers to track the economic health of a nation. These statistics are derived quarterly and are aggregated at a regional or national level (see **Case study 2**). While necessary for understanding the medium- and long-term consequences of a pandemic, they are not sufficient to recognise and act on real-time changes in economic activity, or to identify and evaluate where local policy intervention, such as economic or social support, is needed and effective. There are alternative ways to understand economic activity, including consumer behaviour and payment data, but they require access to, and analysis of, corporately held data.

1. Efficiency:

The right people need to access and analyse the right data at the right time

When responding to a global threat, including a novel pathogen, it is logical that those who can perform the analysis necessary to developing the right insights should gain access to the right data. In ensuring such access, of course, there are trade-offs. However, if these insights form new tools, approaches, and playbooks ensure rigorous privacy protections and are made available global public goods, the trade-off is most likely to be judged worthwhile.

1.1 The right data

It is a critical point that the 'right data' depends upon the specific questions being asked; data for data's sake is not a helpful approach, philosophically or pragmatically. The collection and coding of accurate data is enabled when there is a specific question to be answered; for instance: What is the origin of this virus? How does it spread? What harm does it cause? Broadly, in this pandemic, the data will include traditional scientific and healthcare data, such as the genome sequence of the virus, current and trending test-positivity rates, past and current hospitalisation numbers, nature of treatment administered, and outcomes. Importantly – though this is frequently underappreciated – the right data will also include data collected for other, non-health, primary purposes, but which has clear secondary uses in guiding public-health decisions. This includes, for example, human mobility data from telecommunication or mobile-map companies (see **Case study 1**), household economic behaviour (to inform recovery efforts) via financial transactions recorded by companies, consumer indices on agricultural prices, or levels of viral RNA in local sewerage samples (data from industry to help identify nascent health threats).

In a world where many countries do not have the necessary civil registration and vital statistic (CVRS) systems, there are obvious obstacles to accurate data gathering.[3] Specifically for COVID-19, case reports suffer from ascertainment bias due to insufficient testing, as well as changes in testing intensity and techniques for a given population over time. Similarly, not all data is ready for analysis.[4]

This is also true for non-health data. Although governments keep track of their own policy responses to COVID-19, policies per nation are not available globally. A tremendous volunteer effort led by the Blavatnik School of Government's Coronavirus Government Response Tracker aims to track and compare policy responses around the world, rigorously and consistently.[5] Though crucial, these data-collection efforts are not enough. While they provide a database of what governments are trying to do, they do not include information on adherence, compliance or impact across a range of epidemiological, social and economic outcomes. As a result, we cannot ascertain causal relationships between policy and outcome.

1.2 The right people

As illustrated in Case study 1, understanding the movement of people between cities was critical to understanding the spread of disease in the early stages of the SARS-CoV-2 outbreak. However, the 'right' people – data analysts, data translators and leaders – could not identify how or where to get timeous access to the data. This is not a new problem; data analysts report spending 30-40% of their time looking for the data they need.[6] Public databases remain disparate, fragmented and poorly publicised. These inefficiencies are multiplied in databases held by the private sector and individual academics. Among the inefficiencies are:

1. Information asymmetries. Analysts and researchers are not commonly aware of the richness and granularity of corporately held data, or that this data can be made publicly available via application (with or without a fee). Similarly, organisations are not typically aware of who might use their data and for what purpose, particularly when the primary purpose of the initial data collection is not to improve health outcomes.

2. Access permissions. We have found that researchers typically rely on bespoke agreements or personal connections to gain permission to access data. Standard data-use agreements take time and legal resource to negotiate. Simplified, non-standard agreements are sometimes employed, but these have the unfortunate side effect of creating confusion about how to license the insights, and from whom. Currently, every individual data source must determine the appropriate mechanism for their company to facilitate access, while abiding by local laws.

3. Payment barriers. Due to market incentives, data is considered highly valuable and typically operates on a fee-for-service model. Historically, a lack of incentives as well as licensing issues have resulted in most organisations remaining hesitant to release their data. This can mean that many analysts and researchers who have the potential to add significant value are unable to access expensive or non-available data sources. Pre-existing agreements, prices or access contracts, even for limited use, could speed up access to data under specific emergency circumstances.

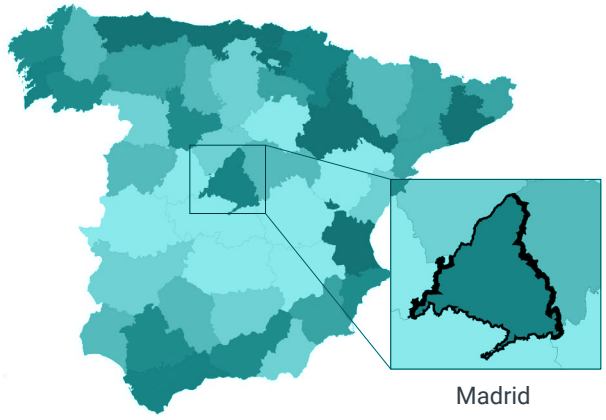
1.3 The right time

Ultimately, the faster data can be shared and analysed and the resulting insights acted upon, the faster we will see improvements in response to, and recovery from, outbreaks, epidemics and pandemics. Today, even if the challenges mentioned above are met, an analyst or decision maker typically gains access after some delay, reducing the efficacy of any response which may follow. A common belief is that these challenges are legal or regulatory, and thus intractable without radical system change. This is misleading and unnecessarily pessimistic. It is now quite clear that this is not an insurmountable challenge, as can be seen in the different data-access arrangements in various countries operating under the same European law.

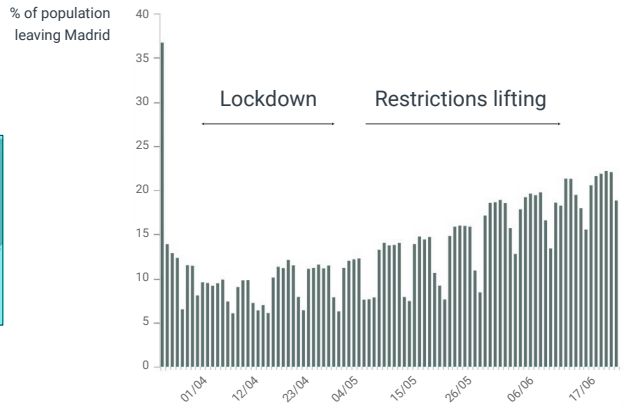
As an example of success, Banco Bilbao Vizcaya Argentaria (BBVA), the University of

Cambridge, Imperial College London, and the University of Edinburgh have used big-data technologies to examine the impact of COVID-19 on Spanish consumption behaviour. To do this, anonymous transactional data yielded a real-time reflection of reactions to the crisis and the economic-policy response, which in turn was used to inform decision making (1.4 billion card or BBVA point-of-sale transactions since 2019 were analysed).[7] Similarly, the Instituto Nacional de Estadística, through collaboration with Orange, Telefónica and Vodafone, has made mobility data from more than 80% of mobile phones available, providing daily point-to-point movement data in fine localised regions (down to a 3km square radius) (see Figure 3).[8] In many countries, neither financial transaction data nor highly granular, real-time mobility data are available to analysts and researchers – certainly not publicly, and most likely not privately – even though these data sets exist. As discussed in Chapter 2, access can be allowed to these data sources in a manner that preserves both privacy and security. Ensuring access to the right data for the right people at the right time does not mean compromising on data rights and liberties, as many sources of useful data are not identifiable.

Distribution of population mobility per province across Spain on 19 June 2020



Tracking of population mobility in Madrid province pre-lockdown to post-restrictions (Apr-Jun)



Percentage of population movement

9.46 - 12.91	15.23 - 16.47	17.27 - 18.08	18.83 - 23.36
12.92 - 15.22	16.48 - 17.26	18.09 - 18.82	

Figure 3 – An INE-led collaboration spanning Spain's leading telecoms providers yielded a rich human-mobility data set, updated daily, allowing decision makers to track movements and test the effectiveness of policies and adherence to mandates.

Case study 1:

Mobility data to understand the spread of Covid-19

There were many unknowns when SARS-CoV-2 was first identified. For example, what was the genetic sequence? How did it spread? Did human-to-human transmission occur? What therapeutics worked? Was immunity from a vaccine possible? Answering these questions required the right people getting access to, and being able to analyse, the right data at the right time. In response to some of these questions, a multinational technology conglomerate collaborated with an academic partner to determine first the pattern of spread in a particular country, and then global spread dynamics. This collaboration has influenced both the scientific literature and the decisions of public policy makers.

At the outset, the academic partner was familiar with infectious-disease spread and dynamics, having responded to many outbreaks over the last 20 years. Their models were ready, but the data was lacking. To understand spread, flight data – typically publicly available, though delayed – is not enough, as in some countries air travel makes up for only 10-20% of total mobility. The key is road or train data, yet these are rarely available. In this instance, however, pseudonymised location data, collected by technology companies through individual use of particular apps, was used to model the spread of disease. Analysis of this data by researchers informed decisions at a local, regional and international level.

What made this rapid analysis possible in some places, but not others?

The primary factor that enabled success was a robust set of individual relationships between organisations. In effect, strong personal connections created a framework for data access, enabling rapid legal arrangements, which ensured subsequent access. Three elements of this relationship are worth emphasising:

Rapid communication. Informal communication via social messaging platforms was used to make initial arrangements. This also ensured that the data was provided in a format that was readily usable, including the right fields for epidemiological analysis, and that changes could be made if necessary. This was distinct from the processes of other technology companies that have made data available, but without the rapid and iterated feedback on whether the analyst found the information useful.

Legal arrangements. The close working relationship ensured that the right legal contract was created. Examples from other jurisdictions show that some organisations have developed emergency, limited-use and time-bound contracts in order to shorten the wait for access. However, this has created its own challenges, as many analysts – and the legal officers at the data-processing organisations – are unfamiliar with the terms of these contracts.

Data remained private. The data was made available from the multinational technology conglomerate on the proviso that it would not be shared further; in effect, it remained private.

The impact of analysis conducted on a data source is only as good as the engagement with decision makers, leading to real-world outcomes; otherwise it is a speculative modelling exercise. Indeed, this analysis informed top-level decisions made by the most senior leaders in the country. We should celebrate this outcome, while equally recognising the limitations of this approach. First, although this analysis enabled the creation of information, greater insights might have been derived had access been granted to more analysts. Second, an inability to share the data further creates challenges for transparency, reproducibility and validation.

Case study 2:

Real-time transaction data to assess policy decisions and economic outcomes

Financial transactions made by households and firms can be used to track the impact of virus transmission and government decisions such as social-distancing policies. These data sources include financial apps on mobile phones, payment-system operators and banks. Financial transaction data has several advantages:

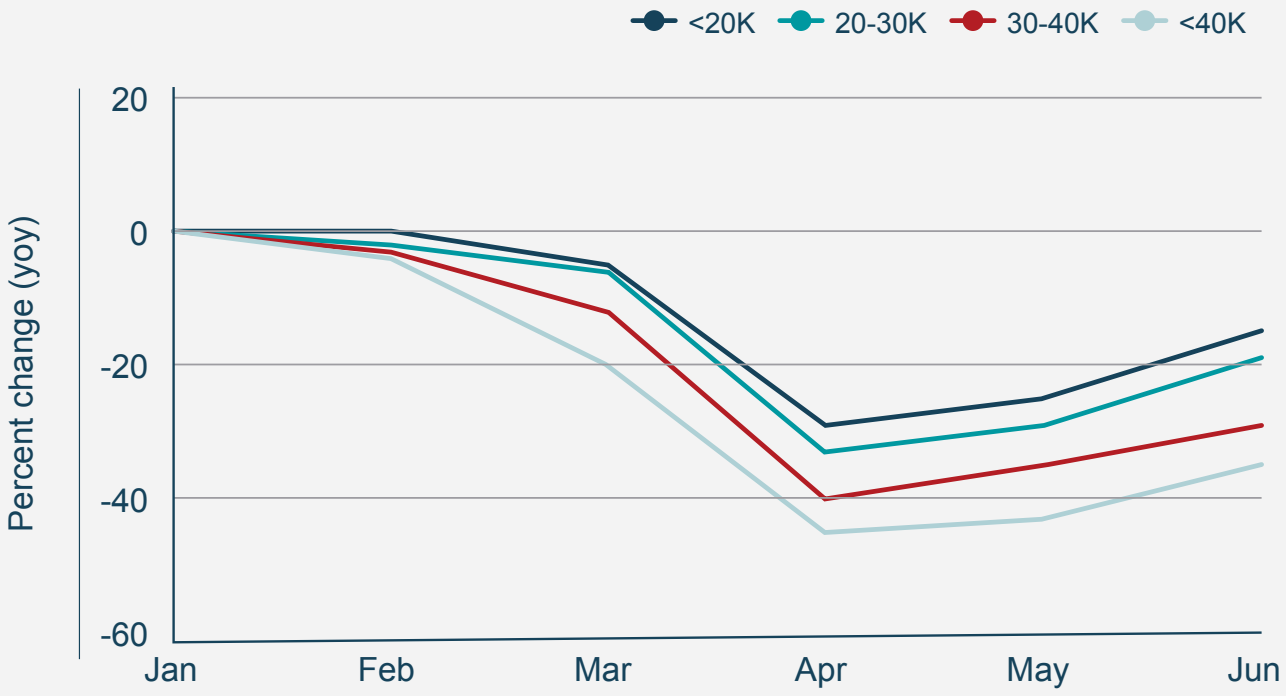
Timeliness. Unlike official national statistics, which can take months to compile, a financial transaction is near real-time. In a crisis, the data can be analysed to provide information which supports evidence-based, timely policy responses.

Granularity. Financial transaction data allows decision makers to know which firms and households are most affected by economic shocks, information that informs complex policy decisions and related trade-offs. An understanding of which sectors and regions are hardest hit by a decision would enable micro-targeting or clustering of economic and social policies.

What does this mean in practice?

Timely, granular data reveals important differences across income groups. High-quality payment data on approximately 15,000 users of a financial app (Money DashBoard) provided evidence that affluent households have cut spending more than others. In the UK, households in the top quartile of the labour income distribution have cut spending (proportionally) the most during lockdown (see **Case study 2**, Figure A).

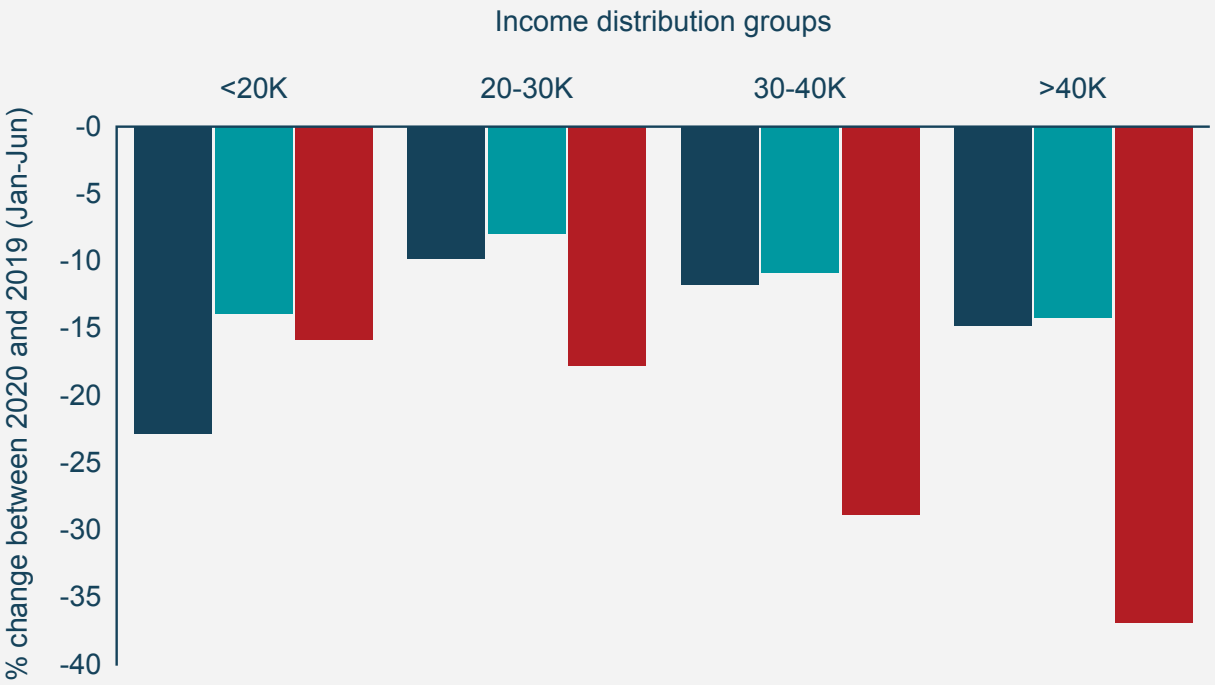
Monthly expenditure by income groups



Case study 2, Figure A – Monthly expenditure by income groups, based on annual after-tax income in 2019

The impact is that more affluent households have increased their savings during the lockdown by reducing non-essential spending on categories such as restaurants, hospitality, recreation, culture and travelling (see Case study 2, Figure B), as a result of restrictive measures and the health and economic crisis more generally.

Implied personal saving rates along the ex-ante income distribution



Case study 2, Figure B – Implied personal savings rates: changes in earnings, income and expenditure, based on annual after-tax income in 2019, for various income groups

Thus, transaction data demonstrates that the decline in consumer spending is driven by richer households, and suggests that any measure to stimulate consumption would have to encourage richer people back into spending.

While this analysis in the UK has provided valuable insights, other countries have done better. Indeed, some countries have transaction data sets with sample sizes that come close to providing a real-time snapshot of spending across the entire economy, and are more representative of consumers across age, occupation and income groups. These data sets could be used to understand very detailed patterns, such as postcode-level spending dynamics and their relation to disease incidence.

2. Effectiveness:

Insights from data analysis should not remain siloed and fragmented

Even when the right people get access to the right data, the insights produced are often not widely shared, limiting our potential to learn rapidly and act across geographies and organisations.

2.1 A lack of information standards

Much has been written about the interoperability of data sources, which is a challenging technical problem. Less has been made of the need to harmonise, validate and then disseminate insights in a timely manner. The World Health Organization (WHO) and other global organisations have made important strides in managing healthcare data. However, to take one example, we still lack many shared definitions, including of COVID-19 cases and deaths. This has meant that, in the context of this pandemic, we lack a standardised, global way of measuring COVID-19 and tracking the pandemic to help guide decision making. The result is great difficulty in sharing and reproducing models and analysis between organisations and geographies. We have thus not built shared knowledge products, codified and shared lessons learnt, or refined hypotheses and approaches in response.

2.2 Information is fragmented

In the UK, at the outset of the pandemic, decision makers did not have access to the accurate, real-time information that was required. They did not know how many ventilators they had in a specific hospital, how many staff were sick and how many others were able to cover shifts, or what PPE was available to those healthcare workers.[9] These are all supply-side concerns, indicating where the system is likely to strain first – whether medical equipment, healthcare infrastructure, or staff health and wellbeing. The same is true for demand: decision makers need information on where the virus is spreading, where it might move next, and how it will affect health and care services.

Insights come not only from healthcare data but also non-health data (Figure 4). In our research, one example is drawn from the fast-moving consumer-goods (FMCG) industry. To understand the impact of national lockdown restrictions, a UK based FMCG company was attempting to understand the change in sales between different intranational regions by analysing the relationship between individual mobility and consumer activity. While useful to validate past commercial decisions, the mobility data to which the company had access was outdated (released six weeks after the fact) and not granular enough (aggregated at too wide a geographic area) to guide real-time, specific insights. Even if data were available in real time, it is not possible to compare and validate this across geographies to learn and improve real-time decision making.

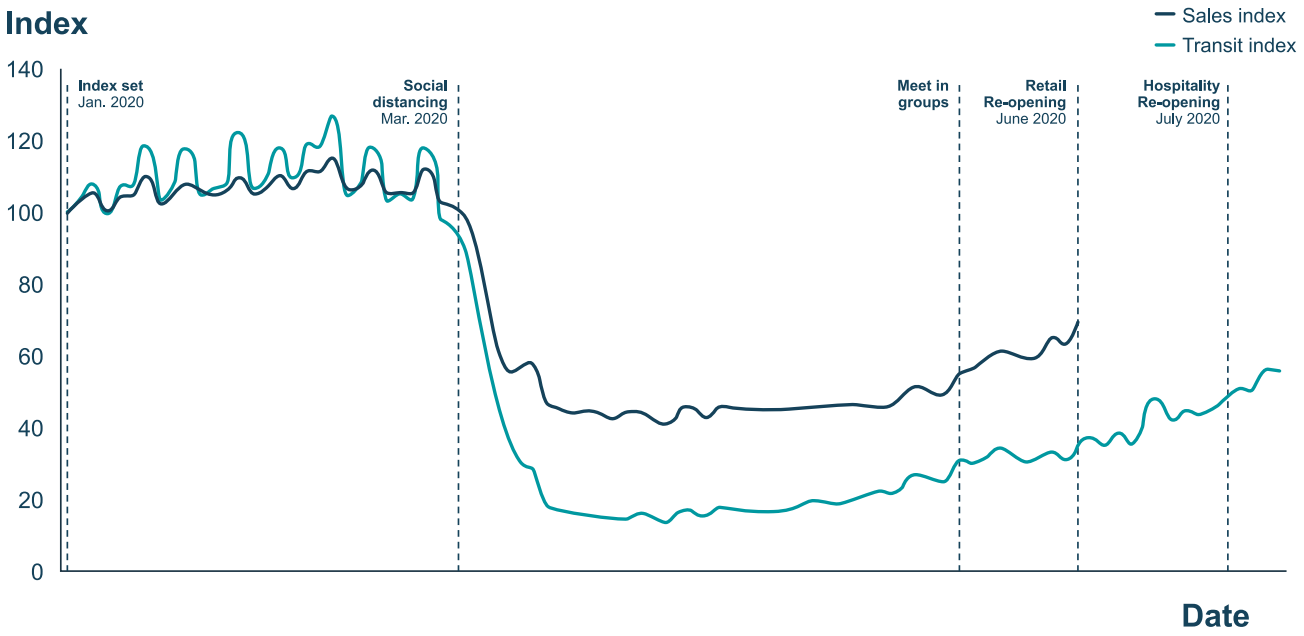


Figure 4 – Non-health data such as consumer preference and human mobility can lead to strong indicators of behaviour in health emergencies; in this example, sales of fast-moving consumer goods in urban settings in the UK track with available mobility data.

2.3 Insights are not shared

Even if complete, timely data analysis were possible, the incentives to share the resulting insights are not aligned. These are not 'geopolitical realities' that plague information-sharing agreements. Rather, they are individual and organisational impediments. For example, individual academics remain beholden to a 'publish or perish' framework, only benefitting when their article is cited in a peer-reviewed publication, which does not prioritise a sharing mentality. Likewise, there is no first-mover advantage for a consumer-goods or mobile company. After all, what do they stand to gain from informing a local or national government about the impact of a particular policy decision? The reality is that organisations are hesitant to release data and information, as their data often underpins their core business. Perhaps this would be acceptable, if it were not for the fact that sharing information is critical to our ability to validate new tools and approaches rapidly.

In short, ineffective use of information is driven by a lack of shared definitions, information is disparate and outdated, and there is a misalignment of incentives to share insights.

3. Sufficiency:

Decision makers should have the information they need

In combination, the inefficient analysis of data and ineffective use of information means that decision makers do not have the information they need to make the best possible decisions in response to evolving, complex challenges. This is certainly a case of 'unknown unknowns': the reality is that decision makers do not know what data and information might be available, or where and how to ask for it. In addition, and as importantly, they do not know what they stand to gain from it.

There remain significant cultural, structural and process disconnections between data collectors and providers, data analysts and decision makers. This is particularly true regarding non-health data and information from the private sector that could be made available to public-sector decision makers. The disconnection

is largely a product of a historical separation between 'the market' and 'the state'. Indeed, we must acknowledge that, when it comes to data, information, and decisions in the public interest, we have historically thought of these entities as strange bedfellows.

Furthermore, leading global organisations have regulations which make it exceedingly difficult to work directly with the private sector. Particularly when it comes to non-health data, a lack of processes and institutional arrangements, and the fact of 'not having done it before', compound the challenges of operating in an emergency. Ultimately, data providers lack the appropriate opportunities to make decision makers aware of the extent to which data and information could assist them. The result is that data generated, collected and analysed, and the resulting information, is not optimised to inform public or individual decisions.

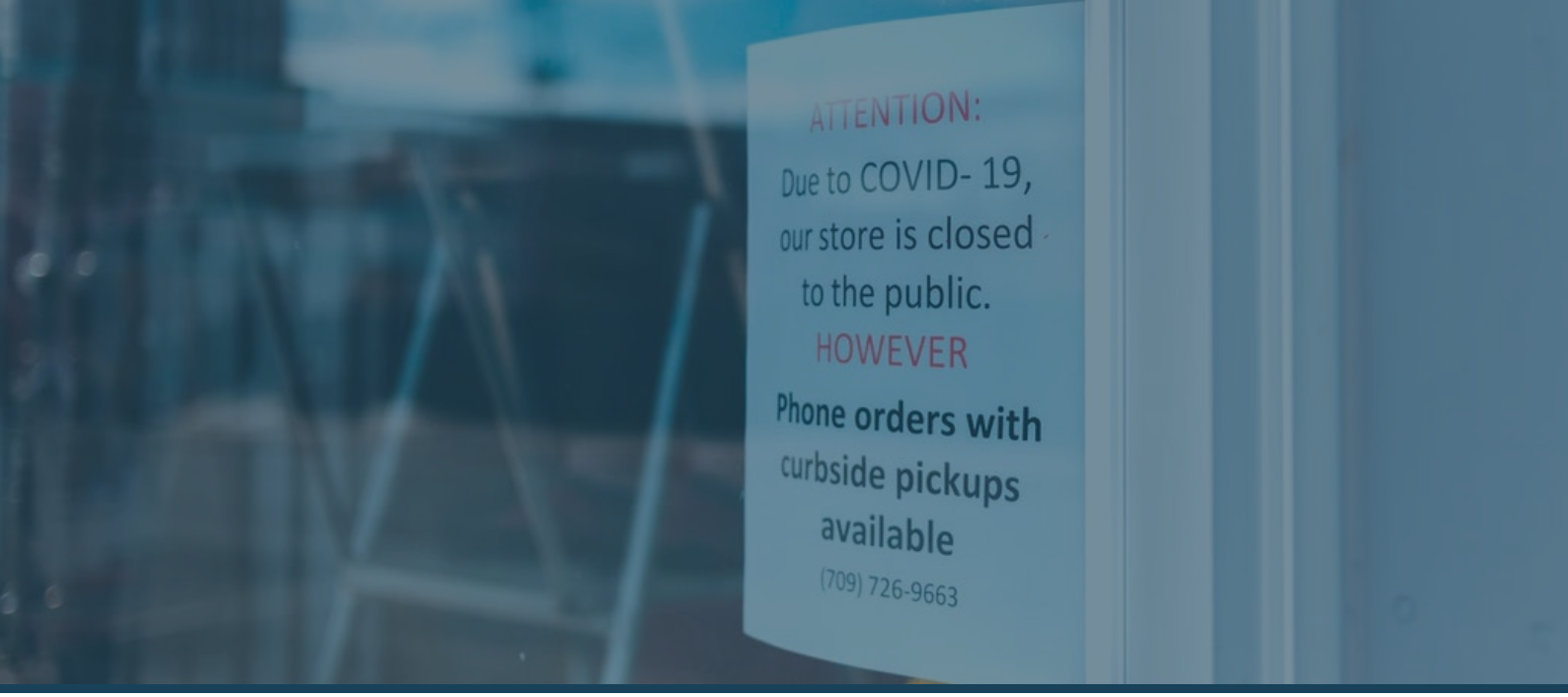
Consider the Facebook Symptom Survey, an opt-in, off-platform survey about COVID-19-like symptoms delivered to a daily sample of the 2.7 billion-plus monthly users of Facebook. (This initiative was started on 6 April 2020 in the United States and on 1 May 2020 globally).[10] Early indications were that the Symptom Surveys could provide a complementary view regarding key, time-sensitive public-health questions about COVID-19 incidence. However, there is a need to increase our understanding of the utility and limitations of symptom-survey data in supplementing public-health outbreak monitoring and response. This validation must be done before such data can be reliably used to inform decisions in the public interest.[11]

Whatever the reasons, we all see that outcomes are universally unsatisfactory. Decision makers do not have the information required to make decisions, and instead must struggle forward in conditions of significant uncertainty. Data providers do not know how their data is being used, and how this might be improved. Data analysts are typically caught in the middle, unsure as to whether they should try to improve their data collection efforts or enhance their analytical methodologies.

Inefficient access to data and ineffective use of information not only limits our ability to make data-driven decisions; it also inhibits our ability to change this process. We lack the required feedback loop to address the root causes of these issues. Addressing this would be collectively advantageous, as improved connections within the data-information-decision process would improve and accelerate decisions made in the public interest.

Finally, decisions made in the public interest must be made with, and communicated to, relevant stakeholders and the public at large. Information must guide decisions, and it is essential to communicate it appropriately and efficiently to those who should know. When there is any element of uncertainty or risk, data and information must be used to explain how and why risk-based decisions are taken.

'Never waste a good crisis' – so the policy-making maxim goes. There is a need and opportunity to do better, but simply understanding the problem is not sufficient; bold leadership is required.



Chapter 2.

Immediate and concrete actions needed

Given the gap between desirable practices and current reality, what is to be done? The challenges set out in the first chapter call for collective leadership across institutions and organisations. This chapter concisely proposes plans for high-impact, immediate actions, while the next chapter outlines long-term aspirations.

Early actions involve increased openness and connection. We propose that governments increase their efforts to identify data that can be shared, and employ effective data governance to make this data more easily accessible; that decision makers in the private sector strive to increase the breadth and depth of efficiently accessible data; and that data providers and decision makers enter into closer communication.

1. Accelerate efforts to make better use of publicly available data

Properly designed, data governance can become a source of value creation, through efficient analysis, effective sharing of insights, and better decisions. Currently, too many data sources are scattered and disparate. From health (such as the fields of microbiology and viral ecology, or the recording of symptoms, cases and hospital capacity), to socio-economic issues (including consumer behaviour, employment figures and economic activity), to government policy, every new data-analytics effort or digital application requires time spent in data discovery, data ingestion, and data cleaning. There is a clear need to create reusable, sustainable and easy-to-access data assets that drastically reduce the time required for data engineering. Research shows that effective data governance can alleviate these barriers: establishing data dictionaries, creating traceable data lineages, and implementing data-quality controls can improve productivity and performance significantly.[12]

The Trinity Challenge data catalogue is a step in this direction, encompassing health and non-health data from public and private sources.[13] The aim is to enable analysts and researchers to avoid wasting time in the pursuit of data. Ultimately, this catalogue will provide information on accessibility by region and type of organisation; history of data, including source, time of collection, and last update; features of the data, including size, structure, format, keywords and scope; and, when relevant, how to get access to content. The next step would be to create data ‘sandboxes’

that allow researchers to share models quickly across borders, whether they are geographic, organisational or sectoral. In addition, increasing efficiencies in the analysis of data, supporting more effective information and improved decisions, would lead to better health, economic and social outcomes. There are similar pre-pandemic national attempts; Health Data Research UK has begun efforts to build public understanding about what data is available, where it can be accessed, and what common legal protocols for access, use and security apply.[14] The International COVID-19 Data Research Alliance and Workbench is an international effort with a similar design.[15]

Leaders can and must:

1. Recognise the importance of signposting which data sources their organisations can make publicly available, and which they can make privately available by application, ensuring rigorous privacy protections.

2. Demonstrate a willingness to share the information from these data sources across organisational and sectoral boundaries.

3. Invest in the curation and maintenance of effective data governance – such as a data catalogue – that is in the public interest.

This effort could be further extended by a different approach to how we value publicly available ('open source') data, and better data documentation and common standards.

2. Strive to extend the breadth and depth of private data that is efficiently accessible

The value of efficient data governance will be amplified as data becomes deeper and broader, whether it is in the public or private sphere, regarding health or other topics. When lives and livelihoods are at risk, regulatory excuses are not valid; if one company is legally able to do something, then all analogous companies operating under the same law should be able to do the same. The decision not to do so is made by an individual organisation and the people in it.

Similarly, technical excuses simply do not hold water; there are multiple privacy-preserving and security-enhancing technologies, such as Microsoft's differential privacy platform, which unlock data while safeguarding privacy, or solutions that have long been used in commercially competitive industries like civil aviation.[16] Making the right data available at the right time will provide invaluable insights from the massive amounts of data that remain locked in corporate and functional silos, as illustrated in **Case study 3**.

Case study 3:

Using differential privacy to protect individuals[17] and enable insights from data at scale

Private, safe and secure data from technology companies can underpin crucial scientific, policy and economic decisions.

For example, with increasing emphasis on public-health strategies, like social-distancing measures, to slow the rate of transmission, public-health officials have sought the type of aggregated, anonymised insights used in products such as Google Maps to help make critical decisions to combat COVID-19. Specifically, Community Mobility Reports provide insights into what has changed in response to policies aimed at combatting COVID-19. The reports chart movement trends over time by geography, and across categories such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces and residential areas.

The ability to share these insights relies upon anonymisation technology used in Google products every day to keep individuals' activity data private and secure. Differential privacy enables the generation of insights while ensuring an individual's data cannot be re-identified.

By incorporating aggregating data, adding noise and hiding data points for which there are few users, differential privacy provides insight into how busy a particular location was, and for how long, without identifying individual users. Since April 2020, when Google first began publishing the COVID-19 Community Mobility Reports,[18] the reports have been downloaded over 16 million times. They are now updated three times a week in 64 languages, with localised insights covering 12,000 regions, cities and counties in 135 countries.

Efforts to collate and analyse disparate, fragmented data sources are not new. For example, in the public sector, we see these efforts in the context of peace and conflict monitoring (led by the Center for Advanced Defense Studies), and in the private sector, in the context of aviation safety (led by Airbus and Skywise). An endeavour that combines the strengthening of data ecosystems, the generation and recognition of insights, and the research and institutional cooperation to act on these insights is the necessary antidote to a global health system which is stuck in its mindset, methodology and membership.

Leaders can and must:

1. Articulate the process through which data from their organisations will be made available, taking inspiration from ambitious and successful practices in comparable organisations.
2. Identify the people (typically legal) in their organisations responsible for making data available.
3. Evaluate the reasons and motivations for keeping data separate and siloed, while maintaining the expectation of efficient data governance and friction-free, effective use of information.

A possible approach to strengthening this effort would be to reconsider how we value corporately held data, and create shared standards and principles for the governance of, and access to, data in a health emergency.

3. Bring data providers and decision makers into closer contact

The benefits of better using public data and increasing access to private data will

become clear if decision makers, analysts, and data generators spend time in regular discussion. As it stands, the feedback loop to move 'unknown unknowns' to 'known unknowns', and ultimately to 'known knowns', is limited. Admittedly, it is challenging to address inefficient and ineffective decisions arising from ineffective sharing of information and inefficient analysis of data. However, the task is not impossible.

Leaders can and must:

1. Broadcast the choices they make, and how they understand the criteria and indicators for choosing one course of action over another.

2. Create opportunities for data generators, analysts and decision makers to engage in regular, high-quality communication about the decisions needed, and the information required to make those decisions; and to determine what data analysis might provide that information.

3. Recognise and celebrate data generators, analysts and decision makers across sectors and organisations as equal partners in responding to complex, evolving challenges.

Creating a tighter feedback loop would be accelerated by a transparent understanding of the data and information needed to make a particular decision, a new approach to how we value public and private data, and common standards and principles for governing and accessing data in a health emergency.

Chapter 3.

How leaders can act with the ideal system in mind

The three immediate actions outlined above would be high-value, no-regret moves by leaders across the private, public and social sectors. Nonetheless, taking these steps is not always easy. System change calls for time, energy and concerted effort, and often comes about only after the components have already begun to move in a particular direction and behave in a specific way. Improving data-driven decisions in response to health emergencies is fundamentally a leadership challenge, not a legal or regulatory one. Likewise, technology is neither a panacea nor a magic bullet. While technological innovation can and will help improve data-driven decision making, it is not necessary, as we can do much better with the data and information we already have available; nor is it sufficient, as investments in technology alone will not help.

The suggested system end-states that are outlined below are not definitively prescriptive. Rather, they seek to demonstrate how leaders might view their ambitious longer-term aims as they take specific, immediate actions.

1. Leading the discussion for a data-governance and access protocol when responding to a health emergency

Leaders need to initiate discussion about effective data governance and access while preparing for and responding to a health emergency. To reiterate, there is no common governance protocol for the use of data – both health and non-health – to improve decision making in a health emergency. Accelerating efforts to make better use of publicly available data and to make more corporately held data accessible would be supported by broader efforts to achieve effective data governance. During a health emergency, data must be accessible in a way that promotes openness, prioritises usability and empowers its users in the search for solutions, while still maintaining privacy and security – and any trade-offs need to be made clear to everyone.

While there are separate protocols and standards for how governments, funders, not-for-profits and civil society relate to digital tools, there is no shared protocol across sectors for how that data could be generated, collected and analysed, or for how the resulting information could be used. This is despite good progress made in other areas of digital health and technology. For example, the Principles for

Digital Development, which were developed in 2017, provide a set of living-guidance principles intended to help practitioners succeed in applying digital technologies to development programmes. Similarly, the Principles of Donor Alignment for Digital Health – which are intended to complement existing development commitments on aid and sustainable development – were signed in 2018.[6] WHO guidelines around point-of-care digital health interventions have also recently been released, including a roadmap for policymakers to introduce and scale digital-health interventions in support of population-health outcomes.[7]

Actors need to answer this question: under what arrangements can which groups of people access data for what purpose, and how is that information to be validated against comparable systems in different locations, and disseminated to inform decisions? Ultimately, an agreement on common standards and principles must cover the collection, aggregation and use of data, including details of a continuous system for reporting data. There is a need for shared terminology regarding what constitutes an item of 'data'; and, particularly when it comes to corporately held data, a need for shared standards about who has access, under what conditions, and for what purposes. These are preconditions for achieving a shared, common agreement on how data can be accessed and information used to improve decision making in a health emergency, enhancing health, economic and social outcomes.

The challenge is at a scale that is likely to require an international agreement for the use of data and information in responding to health emergencies. Alternatively, modifications might be made to existing frameworks, such as the International Health Regulations. Both processes require national governments to lead nationally and work globally through multilateral institutions. While this will undoubtedly be a protracted process, the benefit would be significant. Efficiencies would be found, as analysts would know under what conditions and for what purposes they can access data, and data providers could follow a common process. Effectiveness would increase as insights could be compared and validated across organisations, sectors and geographies.

As a start, leaders must realise and execute their power to begin a conversation about the common standards that are needed to respond better to health emergencies. Leaders can start now by engaging with counterparts at other organisations about common practices of data governance and access in response to health emergencies.

2. Shifting the paradigm from 'the value of data' to 'the value of outcomes'

There is broad consensus that the world would benefit from analysis that crosses disease, regional and institutional boundaries. Yet, as outlined, we lack both the incentives and structures to bring together disparate public-data assets and enable the private sector to contribute to a data ecosystem where we can create, store and validate analysis and models.

Of course, an effort to find common standards and principles for how data ought to be accessed and used in a health emergency must not occur in a vacuum, and must not ignore the opportunity cost of alternative choices. In providing access to corporately held data, there are trade-offs; this data may be used for commercial gain. However, providing access for analysis may provide insights which form new tools, approaches and playbooks to guide better decisions. The outcomes of these decisions, particularly in relation to health emergencies, have benefits for all.

Data collection, cleaning and storage is not cheap. Take the PPE example above; there is a need to align incentives between the suppliers and consumers of the goods, and the data generators (front-line staff), processors (analysts and researchers) and decision makers who then make decisions

about purchasing, distribution and use. The benefit of tracking PPE is not always seen by the person collecting the data, and hence there must be other forms of incentives or investments to ensure the right data is made available to the right people at the right time.

We need a framework for understanding and measuring these benefits, thus gaining a more nuanced understanding of how to price and value data. Instead of a paradigm where we are concerned about the value of data per se, we need to move to a paradigm where we recognise the value of outcomes and advance the idea of 'data philanthropy'. Different circumstances necessitate different operating models, and health emergencies require a new way of conceptualising and valuing data.

3. Shaping the development of a playbook for decision-makers

Finally, with regard to common governance and access arrangements, and a new paradigm of valuing data, these will only be beneficial insofar as they improve data-driven decisions. When it comes to enhancing the efficient analysis of data and effective use of information, we are ultimately interested in the destination, not only the journey.

Recognising that local circumstances differ, and that decisions are ultimately a balance of information, resources available and political judgement (both governmental and corporate), there is a need for a common playbook for decision makers to identify, respond to and recover from health emergencies. This playbook would include:

- An appropriate ordering of the questions that need answering, including: What is currently happening? What are my options to respond? What is the expected impact of any specific decision? How do I monitor the actual impact? At what stage do I evaluate and refine a specific decision?
- A catalogue of the different sources of data – both health and non-health – and information required to make these decisions
- An outline of which organisations can provide the data, and which analysts will generate the insights.

These suggestions are by no means exhaustive. Instead, they are examples of ways in which leaders can contribute to systems, processes and infrastructural arrangements that allow us not only to improve our response and recovery during this crisis, but to do better next time.

Leaders need to take action on six shifts in data governance, to better protect people against health emergencies

	From	To	Rationale
	Scattered, disparate publicly available data	Reusable, sustainable, and easy-to-access data assets	Improve productivity and performance
	Minimal privately held data sources available	The right data is made available to the right people at the right time	Enhance the generation, validation and dissemination of insights
	Fragmented decision-making process	Increased connectivity between data generators, analysts and researchers, and decision makers	Enable efficient and effective feedback loop with learning and improvement in real time
	Lack of shared access protocols	Common understanding about who can access what for what purpose	Accelerate effort to make better use of public and privately held data
	Silos that focus on the value of data	Systems that focus on the value of outcomes	Enable incentives and investments to strengthen data ecosystem
	Data hoarding where time, effort and capital are invested without return	Purpose-driven, accurate, real-time data collection	Sensitive and sophisticated data collection and use of data to improve decisions and outcomes

Figure 5 – Leaders can do six things to ensure that data-driven decisions are taken to protect people against health emergencies.

Conclusion

The evolving global health agenda provides opportunities to address inefficient analysis of data, ineffective sharing of information, and inadequate decisions. For example, the UK Government's plans for global health security during its presidency of the G7 - specifically the fourth objective, 'agreeing global protocols for future health emergencies' – provide a forum to engage.

The perspectives in this paper are not a comprehensive answer to the challenges of making data-driven decisions for the public benefit in an emergency. Instead, the authors have attempted to provide a sense of the definition, nuance and boundaries of the challenges we face amidst inefficient data analysis, ineffective sharing of information, and ultimately inefficient and ineffective decisions. In addition, this paper has asserted immediate actions that leaders can take now, for the COVID-19 pandemic and for future health emergencies. Finally, the paper has articulated the features and functions of a future data and analytics ecosystem, maintaining that for any future pandemic we will need an industrialised data and analytics platform to integrate all sources of local data, make accurate forecasts to inform policy decisions, learn rapidly from effective local interventions, and forecast scenarios on a broader scale. All of this calls for learning in real time and reacting in a much more granular manner.

The actions proposed for leaders provide a plan for organisations and individuals committed to ensuring better protection against health emergencies. The Trinity Challenge, an initiative led by a set of prominent organisations across the private, academic and social sectors aiming to find solutions that better protect one billion more people from future health emergencies, is a pilot in this form of collaboration and joint leadership. These challenges cannot be overcome by one group or actor alone; we must come together in concert to accomplish what we cannot individually.

Endnotes:

[1] *Nehru, J. (1948). The Unity of India: Collected Writings 1937–1940, London: Lindsay Drummond.*

[2] *Oliver, N. et al. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Science Advances, 6(23). https://advances.sciencemag.org/content/6/23/eabc0764.*

Grantz, K.H., et al. (2020). The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. Nature Communications, 11(4961). https://www.nature.com/articles/s41467-020-18190-5.

[3] *Mikkelson, L., et al. (2015) A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. The Lancet, 386(10001), 1395–1406 https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)60171-4/fulltext*

[4] *Lawrence, N. D. (2017). Data Readiness Levels. Cornell University. https://arxiv.org/abs/1705.02245*

[5] *Blavatnik School of Government. Coronavirus Government Response Tracker, University of Oxford. https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker*

[6] *Grande, D., Machado, J., Petzold, B., & Roth, M. (2020). Reducing data costs without jeopardizing growth. McKinsey Digital. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/reducing-data-costs-without-jeopardizing-growth*

[7] *Carvalho, V. M. et al. (2020). Tracking the COVID-19 crisis with high-resolution transaction data. BBVA Research. https://www.bbvaresearch.com/publicaciones/seguimiento-de-la-crisis-del-covid-19-con-datos-de-transaccion-de-alta-resolucion/*

[8] *Evolution of mobility by geographical area during the alarm state by COVID-19. https://www.ine.es/en/covid/covid_movilidad_en.htm?L=1*

[9] *Gould, M., Joshi, I. & Tang, M. (2020). The power of data in a pandemic. Technology in the NHS. https://healthtech.blog.gov.uk/2020/03/28/the-power-of-data-in-a-pandemic/*

[10] *Privacy is paramount in the Symptom Survey; Facebook invites Facebook App users to take the survey conducted by partner universities via an invitation at the top of their newsfeeds and does not receive individual responses.*

[11] *Validation is also needed for self-reported behaviour data like mask wearing, for mobility data, or for Google's recent symptom-search data.*

[13] *The Trinity Challenge. Data Sources. https://www.hdruk.ac.uk/covid-19/international-covid-19-data-alliance/*

[14] *Health Data Research UK. Our Hubs. https://www.hdruk.ac.uk/covid-19/international-covid-19-data-alliance/*

[15] *International COVID-19 Data Alliance (ICODA). https://www.hdruk.ac.uk/covid-19/international-covid-19-data-alliance/*

[16] *Kahan, J. (2020). New differential privacy platform co-developed with Harvard's OpenDP unlocks data while safeguarding privacy. Microsoft Blog. https://blogs.microsoft.com/on-the-issues/2020/06/24/differential-privacy-harvard-opendp/; Skywise. Unleashing the potential of data. Airbus. https://www.airbus.com/aircraft/support-services/skywise.html*

[17] *YouTube. Helping developers and organizations use differential privacy. https://www.youtube.com/watch?app=desktop&v=FfAdemDkLsc&feature=youtu.be&ab_channel=Google*

[18] *Google. COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/*

[19] *Yokoyama, J. (2020). Closing the data divide: the need for open data. Microsoft Blog. https://blogs.microsoft.com/on-the-issues/2020/04/21/open-data-campaign-divide/*

The COVID-19 pandemic continues to wreak havoc on our lives and livelihoods. Many governments, institutions, businesses and communities across the world have taken bold and decisive action to protect lives and mitigate the economic impact of the pandemic – yet the events of the past year have revealed profound gaps in humanity's preparedness for health crises. It is clear that all actors can and should work together to do better next time.

Authored by:

Connor Rochford (The Trinity Challenge), Sally Davies (The Trinity Challenge), Lucinda Scharff (Google Health, UK), Dominic King (Google Health, UK), Mitch Cuddihy (McKinsey & Company), Lars Hartenstein (McKinsey & Company), Hemant Ahlawat (McKinsey & Company), Kristin-Anne Rutter (McKinsey & Company), Kevin Sneader (McKinsey & Company), Gabriel Leung (HKUMed), Joseph Wu (HKUMed), Anna Vignoles (University of Cambridge), Samuel Scarpino (Northeastern University), David Luzzi (Northeastern University), Alexander Ng (Tencent), Ling Ge (Tencent), John Godfrey (Legal and General), Kerrigan Procter (Legal and General), Ara Darzi (Imperial College London), Emile Stipp (Discovery)

Edited by:

Colin Douglas and Greg Fried (Douglas Knowledge Partners)

Design by:

Simon Säll and Anthony Browne (Formation)

THE TRINITY CHALLENGE