Building Real-Time Data Infrastructure Into the Heart of Public Health

HENRY FINGERHUT

TONY BLAIR
INSTITUTE
FOR GLOBAL
CHANGE

Contents

Executive Summary

How does the UK's health-data infrastructure affect the ability of government to make precision-based policy decisions? And what lessons does it hold for the future of high-visibility, data-driven decision-making in the public sector? We explore the following key points in this paper:

- Data infrastructure is the core foundation for evidence-based decision-making and is essential to the Covid-19 response.

- The UK government has come a long way in bringing together novel and existing data sets from across institutions and disciplines, including non-traditional data sources, with Health Data Research UK (HDR UK) playing a coordinating role.

- Despite these advances, the government could do more to join up granular civic and health data to inform targeted interventions at the local level to support the most at-risk communities. Considering the privacy concerns around data that may arise as a result, government should clearly and transparently communicate the trade-offs.

- Policymakers should also create novel platforms, such as an online research community, to engage researchers – particularly experts – in near real-time analysis of non-traditional data. This would reduce the lag between epidemiological phenomena and government decision-making in terms of pandemic response.

The Covid-19 pandemic demonstrates the opportunities and challenges of public-sector innovation. Public-sector bodies in the UK and abroad have designed and implemented real-time data infrastructure in real-time, demonstrating the often-underappreciated agility of government at its best. But the false starts and mixed messages that have in part characterised the pandemic response – in turn, adversely affecting public confidence, compliance and outcomes – indicate the challenges of designing data infrastructure in real-time and at scale while translating data into policy decisions.

So, what can we learn from the pandemic about public-health data infrastructure, evidence-based decision-making and public-sector innovation at large?

Infrastructure for Data-Driven Decision-Making

Over the past year, governments across the globe have taken different approaches to implementing data infrastructure and analytical capacity at differing degrees of centralisation, enabling them to precisely target the Covid-19 response to the needs of specific groups and localities.

As we have previously set out, good data infrastructure underlies the situational awareness necessary for precision policy decisions. This targeted approach represents a third way between a broad-brush, population-level response that is overly permissive (thus allowing the virus to spread and burdening health systems) and one that is overly restrictive (thus burdening the economy and constraining civil liberties).

The UK government, in partnership with private-sector, academic and other public-sector actors, has used a wide range of data infrastructure to respond to the pandemic, in many cases building systems from scratch in short order. Examples are given in Figure 1.

Figure 1 – Data infrastructure established for the UK's pandemic response

Database	Description	Sources/Stakeholders
REACT (Real-time Assessment of Community Transmission)	Epidemiological: real-time data to track community transmission via home testing	New. Includes Imperial College London, Ipsos MORI and Imperial College Healthcare NHS Trust
National Pathology Exchange (NPEx)	Operational: enables professionals within NHS trusts to route pathology tests to available labs and track patients' pathology results	Existing but has expanded its participation mandate for Covid-19. Includes

		NHS trusts and pathology labs
HDR UK	Coordination: relevant data sets and research coordinated across institutions and disciplines. Fortnightly reports to SAGE	Existing but dedicated programmes set up for Covid-19. Includes academic institutions, NHS and SAGE
NHS Digital Dashboards	Epidemiological, operational and public communications: disease progression, comorbidity, shielding information and operational capacity, aggregated to local, CCG and national levels	New. Includes NHS and ONS
GOV.uk Dashboards	Epidemiological, operational, civic services and public communications: cross-cutting dashboards on testing, cases, vaccinations, health-system operations and deaths, aggregated to local authority, NHS trust and national levels	New. Includes NHS, public-health bodies and government
ZOE	Epidemiological: real-time infection estimates based on opt-in participation from ZOE app users, aggregated to regional level	New. Includes ZOE, King's College London and app users
Covid Economic Round Up	Economic: traditional economic measures augmented for Covid-19, including relevant census data, unemployment, consumer and business confidence levels	Existing but expanded by the ONS with Covid-19-specific metrics.

The Successes and Opportunities

Scope of the data: pandemic-response data infrastructure spans the wide-ranging metrics necessary for both government and individuals to take decisions in the pandemic, including epidemiological data, health-system operations, and economic metrics at the local, regional and national levels.

Range of the partners and data sources involved: the UK's coronavirus data infrastructure has emerged through the interactions of public, private and academic sources. They have emerged on an ad hoc basis – as in the case of ZOE's Covid Symptom Study, which was launched in partnership with King's College London during the early stages of the pandemic to engage app users and enable symptom-based tracing before tests were widely available. But the Department of Health and Social Care has also commissioned specific real-time analysis and research programmes such as REACT among academic, private and health-sector actors.

Independently, academic researchers have collaborated in novel ways across disciplines and countries through groups like the Covid-19 Dispersed Volunteer Research Network and the Covid-19 Mobility Data Network to address a pandemic that transcends traditional frontiers. The use of non-traditional data sources – from search queries and Twitter sentiment to smartphone mobility and cell tower data – has enabled these researchers to conduct more cross-cutting, practice-oriented analysis to inform policymaking in the short-term. For example, one research team has developed a method to estimate case prevalence in near real-time via online search queries while simultaneously minimising expected bias in these signals as a result of general public interest (rather than infections). The Scripps Research DETECT study and Stanford Healthcare Innovation Lab both use sensor data from wearable devices and mobile phones to screen for possible Covid-19 cases via changes in heart rate, sleep and activity levels. University of Cambridge researchers have used voice and breath recordings to predict Covid-19, and a similar technology has recently been commercialised at scale. Over the long term, through traditional academic channels, this research contributes to both theory on pandemic progression and the continued development of novel research methods for the use of non-traditional data.

Speed of implementation: data management and visualisation systems have been built from scratch or greatly expanded and interconnected in short order, demonstrating the public sector's undervalued agility in certain contexts. The coordinating role played by HDR UK has facilitated the implementation and joint analysis of data infrastructure across institutions and disciplines, as well as translating data and findings up to SAGE in fortnightly briefings to inform government decisions.

Public communication and data literacy: both the government and NHS digital dashboards have emerged as innovative public communication and transparency tools, conveying the breadth of relevant epidemiological and operational data and visualisations interactively, in real-time and at scales that enable individuals to hold government to account while making informed behavioural decisions. Similarly, these

resources, along with effective data journalism in the media, will help improve data literacy in the general public, engaging people on the complexities of analysis to inform individual decisions. The effects of seven-day moving averages, test sensitivity versus specificity, timeseries lags and complex age or geographic distributions have become routine in media coverage of the pandemic. A large proportion of the population have low numeracy, or difficulty interpreting and making decisions using numerical information, and so are susceptible to cognitive biases as a result, particularly in relation to medical choices. The pandemic alone will not resolve this, but advances in data journalism over the past year have gone a long way towards integrating these concepts into common policy discourse.

Meeting the Challenges of Today and the Future

The design, procurement and use of these systems in the UK have also demonstrated the challenges that governments continue to face in applying data and evidence to policy decisions.

Cost, scope and speed of programmes: given the scale, timing and critical nature of these data-infrastructure and service-management systems, the risks of incurring adverse costs or not securing quality outcomes are understandably real. Yet system design and procurement choices made by the UK and US governments have led to lucrative consulting contracts for unwieldy systems that have since been criticised for not fulfilling their core purposes. Indeed, the UK's Test and Trace system, which built PCR testing and expanded laboratory capacity, has been criticised in a parliamentary study for not having a demonstrable impact on transmission reduction, via contact tracing, and for its overall value for money in light of the non-competitive procurement process used. Similarly, the US Vaccine Administration Management System faced technical challenges of operating at scale across state jurisdictions; though intended to coordinate national vaccine distribution and provision, the system was regularly bypassed by clinics and end users in daily operations.

Real-time data dynamics: real-time decision-making is constrained by the challenges of collecting and analysing real-time data. Time lags are a function of disease mechanics, testing mechanics, data analysis and decision-making. Data represent a "photograph" of conditions at a given time; what's more, different data "moves at different speeds", making it difficult to combine in order to draw policy conclusions. Modelling pandemic dynamics and predicting the effects of different government interventions is also computationally time-intensive although novel techniques have emerged to reduce this limitation through initiatives like Rapid Assistance in Modelling the Pandemic (RAMP). While the government has built a five-week decision-making cycle into the latest lockdown to account for these delays, these dynamics have made it difficult to time interventions such as school closures and hotel quarantine to current epidemiological conditions.

Merging civic and health data: the UK's pandemic response is a systems challenge. Decisions are taken and data collected independently by governmental departments, civic service providers and health-care

organisations at the local, regional and national levels. While government and the public can observe aggregate data from each of these providers, true situational awareness for government and individual decision-makers requires data to be merged and analysed on a more granular basis. Public dashboards may feature trends aggregated separately from community channels (for example, schools) and from health systems (for example, at-risk populations as assessed by Test and Trace), but targeted action at both the government and individual levels requires a granular understanding of how these programmes, interventions and risks interact. When data from various sources does not align, granular data in real-time could help local authorities identify outbreaks and respond. While epidemiologists have worked at an impressive pace to produce scientific knowledge about transmission and disease dynamics, these findings are necessarily retrospective and thus fulfil only a part of real-time decision-making needs.

To date, the HDR UK Innovation Gateway has played an important coordinating role linking relevant genomic, testing, economic, demographic and opinion data sets across institutions in response to Covid-19. The regular HDR UK report to SAGE includes information on the linkages available across data sets and ongoing research initiatives leveraging these resources. Going forward, these important research channels – including not only health but also relevant civic and economic data – should be preserved to facilitate real-time, data-driven decision-making on cross-cutting policy issues.

Recommendations for Today and the Post-Pandemic Era

System Design and Procurement

In the short term, the UK government should commit to both procurement mechanisms and system design specifications that diversify access to contracting, for example, by designing more decentralised or modular systems. Novel approaches to procurement, particularly those that enable SMEs to compete, should be applied in the interest of market access, cost and innovation.

Post-pandemic, the government should strengthen public-sector capacity to build such systems internally. Groups like the Government Digital Service – responsible for the GOV.UK website – and the Open Innovation Team have played a strong role both in building internal digital and data analytics capabilities within government and facilitating procurement among diverse, high-quality vendors. The government should look at extending these initiatives within the health domain.

Communication of Risk and Policy

In the short term, the government should more clearly communicate the decision-making criteria and risk trade-offs they use to make pandemic policy. Though transparent with underlying data and visualisation, the government has been reluctant to specify the criteria supporting key decisions including lockdowns, school closures and quarantine policies. As the public only has access to aggregate data, it is unable to determine whether broad-brush government decisions and false starts (such as the January schools reopening) are based on data limitations, in which case such systems should be prioritised, or whether an intentional policy decision has been made to prioritise certain risks over others. Ambiguous statements and optimistic claims cloud the government's pandemic risk strategy and erode public confidence.

The government has instituted a framework to clearly communicate decision-making criteria to local authorities, essential to a lockstep pandemic response at all levels of government. Communicating these criteria to the public, however, is also critical to improve public confidence in – and compliance with – policies.

Joined-Up Civic and Health Data

Data dashboards are a powerful decision-making tool that enable policymakers and individuals to quickly scan and internalise information from disparate sources in real-time. But often those sources remain siloed, compared only in aggregate when decision-makers need information on granular interactions. Joining up individual-level data from across public services and health-care settings must be done with strict adherence to privacy standards in order to maintain public trust. Stakeholders could certainly argue, as the Institute did early in the pandemic, that privacy is "a price worth paying" compared with an overwhelmed health system and economic shutdown. But government should make explicit and transparent the limits and trade-offs it intends to make so the public understands whether privacy curtailment is a system-design problem or conscious civil liberties sacrifice. With this knowledge, actors could design appropriate, privacy-preserving solutions for schemes such as tracing, individual quarantine or vaccines passports accordingly.

In the short term, building on the recommendation above to clearly communicate any risk-management approach, policymakers should continually reassess the trade-offs between those risks and privacy concerns associated with merging health data and real-time location data, for example, at the individual level across different sources. When policymakers determine the price is indeed worth paying, for instance because hospitalisations are above an acceptable level of risk, they can act accordingly and trade off a certain degree of privacy to protect the health system and lives. But it is only possible to do so in a transparent, accountable manner with continuous measurement of risks and a clear statement of risk profiles.

Policymakers should also actively engage with and support ad hoc research communities applying non-traditional data sources, by co-creating research and public-service design agendas based on these methods. Co-creation ensures that researchers' analysis aligns closely with the immediate needs of policymakers, while facilitating governments to readily implement their findings as policy inputs or as automated public services that prioritise the most at-risk localities. Specific recommendations in this area include:

1. Employing non-traditional data sources to reduce decision lag

Government should directly engage researchers with expertise on the use of non-traditional data for epidemiological estimation to reduce the five-week decision-making cycle imposed by disease and testing dynamics. While these methods would not replace traditional epidemiological studies, they should be used in real-time to track system behaviour and provide early warnings that may impact decision-making. Public Health England now includes this data in its weekly surveillance report at the national level; these trends should also be extended to other non-traditional sources and evaluated at the local level to identify risks early.

2. Supporting hackathons, communities and broader academia-government collaboration

While many ad hoc research groups have formed to respond to the pandemic, they operate largely in silos, unaware of what others are doing and what governments need most. Certain agencies, notably NHS Digital and the Natural Environment Research Council (NERC), sponsored one-off hackathons early in the pandemic but these initiatives have not been continued. Agencies should create mechanisms to mobilise and sustain researchers not only during a development sprint but through the implementation and sustainment of pandemic-related public services. Agencies could, for example, create an online researcher community, posting open questions and requests for public-service proposals – while also allowing for community vetting – along with appropriate access to data to rapidly generate analysis and service designs directly relevant to government needs. Such a programme would democratise public-service design and mobilise a diverse range of highly skilled researchers, designers and data scientists across disciplines to create and vet services.

Post-pandemic, the ongoing shift to the Integrated Care System structure within the NHS, which aims to join up local authorities and NHS trusts to provide comprehensive health and social care, represents an opportunity to more broadly link civic services and healthcare provision at local and national levels. The UK government should consider what data infrastructure is necessary to provide precision public health more broadly, enabling policymakers, civil servants and health practitioners to identify risks – especially ones that cut across community and health factors such as socioeconomic status – and target interventions accordingly.

Conclusion: Working Towards Precision Public Health

The Covid-19 pandemic has demonstrated the susceptibility of the public and health sectors to shocks while simultaneously underlining their underrated ability to innovate at speed. Data infrastructure is the foundation for evidence-based decision-making and precision public health. The pandemic has demonstrated just how important it is to join data from across academic disciplines, levels of government, and both traditional and non-traditional sources to rapidly target public policy and service interventions where they are needed most and to generate reliable theory about the emerging threat in near real-time.

The UK has in place the interconnected data systems and institutions necessary to address cross-cutting health policy challenges such as Covid-19. HDR UK coordination and regular reporting to SAGE are institutional mechanisms enabling this research to inform government decision-making. Given that these structures are in place, the policy challenge becomes one of implementation: building decision support tools, on top of the data systems, which route the right information to policymakers in the right form, conducting analyses using granular civic and health data to inform a range of government policy options, and transparently communicating policy priorities and risk profiles to the public. With the data systems in place, these strategic choices will enable government to design targeted policy interventions and gain the public confidence and compliance necessary to address the pandemic and broader health challenges.

While Covid-19 has demanded and driven many responses from our health-data infrastructure, there is still much more to achieve now and to build on for the future. Both technological and organisational innovations are needed to ensure that governments can source not only the data they need but the diverse skills and perspectives they need as well in order to translate it into effective, responsible policy decisions. The current pandemic is not over but there is an urgent need to keep developing the right mechanisms so that we can deal with health crises and outbreaks when they occur in the future.

FOLLOW US

facebook.com/instituteglobal
twitter.com/instituteGC
instagram.com/institutegc

GENERAL ENQUIRIES

info@institute.global

FIND OUT MORE
INSTITUTE.GLOBAL