


RESEARCH ARTICLE

Responding to the coronavirus disease-2019 pandemic with innovative data use: The role of data challenges

Jamie Danemayer¹ , Andrew Young² , Siobhan Green³, Lydia Ezenwa⁴ and Michael Klein⁵

¹Global Disability Innovation Hub, University College London, London, United Kingdom

²The GovLab, New York University, New York, NY, USA

³DT Global, Washington, DC, USA

⁴AfriLabs, Abuja, Nigeria

⁵Itad, Washington, DC, USA

Corresponding author: Jamie Danemayer; E-mail: jamie.danemayer.21@ucl.ac.uk

Received: 19 January 2022; **Revised:** 11 November 2022; **Accepted:** 07 February 2023

Keywords: COVID-19; data governance; data innovation; low/middle-income countries; responsible data use

Abstract

Innovative, responsible data use is a critical need in the global response to the coronavirus disease-2019 (COVID-19) pandemic. Yet potentially impactful data are often unavailable to those who could utilize it, particularly in data-poor settings, posing a serious barrier to effective pandemic mitigation. Data challenges, a public call-to-action for innovative data use projects, can identify and address these specific barriers. To understand gaps and progress relevant to effective data use in this context, this study thematically analyses three sets of qualitative data focused on/based in low/middle-income countries: (a) a survey of innovators responding to a data challenge, (b) a survey of organizers of data challenges, and (c) a focus group discussion with professionals using COVID-19 data for evidence-based decision-making. Data quality and accessibility and human resources/institutional capacity were frequently reported limitations to effective data use among innovators. New fit-for-purpose tools and the expansion of partnerships were the most frequently noted areas of progress. Discussion participants identified building capacity for external/national actors to understand the needs of local communities can address a lack of partnerships while de-siloing information. A synthesis of themes demonstrated that gaps, progress, and needs commonly identified by these groups are relevant beyond COVID-19, highlighting the importance of a healthy data ecosystem to address emerging threats. This is supported by data holders prioritizing the availability and accessibility of their data without causing harm; funders and policymakers committed to integrating innovations with existing physical, data, and policy infrastructure; and innovators designing sustainable, multi-use solutions based on principles of good data governance.

Policy Significance Statement

A healthy data ecosystem can support an efficient, effective response to the coronavirus disease-2019 (COVID-19) pandemic and future emerging threats. Creating this ecosystem and sustaining it requires investment and awareness from decision- and policymakers at global, national, and regional levels. The challenges arising from data-poor settings, which lack good data governance, availability, and infrastructure, are described in our study by both data innovators and policymakers. Responses to these gaps (including best practices and specific tools) are also introduced. Recommendations specifically for policymakers are provided, emphasizing the importance of integrating innovations with existing infrastructure and supporting bottom-up approaches to enable innovations to thrive in a unique context and continue to serve the community beyond the COVID-19 pandemic.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (coronavirus disease-2019 [COVID-19]) pandemic has affected communities around the world. In both the highest- and lowest-income countries, leaders have had to contend with serious threats to health as well as social and economic well-being. Often in this effort, policymakers and others working in the public interest have turned to data to provide them with the insights they need to make sense of complex and developing situations and achieve their goals (OECD, 2020). In all income contexts, these actors face barriers in accessing useful data and matching the demand for data with its supply. Potentially impactful data, particularly in low/middle-income countries (LMICs), are often siloed, fragmented, nonexistent, or otherwise not accessible to those who could put it to use to help address COVID-19 (Klein, 2021). Data challenges constituting calls-to-action for innovative methods of data use provide a valuable opportunity to better understand these barriers and the role of the data innovation ecosystem in the context of the pandemic, where traditional/existing data sources may be lacking or go unused. This study synthesizes learnings from three distinct datasets: innovator applications to the COVIDaction data challenges, surveys from organizers from similarly-aimed data challenges, and a focus group discussion with professionals who work with COVID-19 data. Thematic and topic analyses were used to analyze these datasets with the aim to identify gaps and barriers to effective data use in responding to the pandemic.

The need for timely, reliable data to mitigate COVID-19, as well as barriers to meeting this need, have been highlighted throughout the pandemic. For example, in an emerging situation, it is essential for governments and ministries to make evidence-based decisions (Harrison and Pardo, 2021) and be considerate of vulnerable populations. The effective communication of research and real-time data is also essential for public education (Röddiger et al., 2021), which affects participation in mitigation interventions and efforts. Mobility data has further been adapted as a proxy to understand individual behavior and geographic movement during the pandemic, also highlighting best practices regarding ethical data re-use (Ågren et al., 2021; Kishore, 2021). However, using data to create public value requires a unified data infrastructure and governance principles. Barriers to equitable data use have arisen during the pandemic (Jagadish et al., 2021), which also include the responsible collection of data with respect to privacy (Redmiles, 2021). These gaps, like many inequities and service delivery system shortcomings, have been exacerbated by the pandemic.

However, a broad understanding of these needs and gaps is not sufficient to identify and address barriers to effective data use at local levels. These bottlenecks and data silos are especially challenging to address in data-poor contexts, specifically because information is lacking in these settings. There is a need to learn from innovators designing in LMICs, organizers supporting these innovations, and the professionals who use COVID-19 data to inform policy and decision-making at local levels.

Further, it is difficult to connect these stakeholder groups and synthesize learnings from the data innovation ecosystem at large, as it is highly fragmented (The GovLab, 2021). Existing research is disparate and often focuses on a specific type of innovation (Ahmed et al., 2020), model (Angarita and Nolte, 2020), or initiative (Gama, 2021). As the pandemic continues to be an emerging threat, and innovators and organizers attempt to rapidly respond to growing needs, even less is understood about the innovation ecosystem's role in the context of COVID-19. Therefore, an essential first step to advance effective data use is organizing what opportunities and data challenges are available to innovators and useful to policymakers to mitigate the pandemic.

#Data4COVID19 repository

Innovators around the world are navigating data challenges to fill data gaps and support pandemic response. In its #Data4COVID19 repository (The GovLab, 2021), the GovLab collected over 230 examples of ways in which policymakers have harnessed data for pandemic response between March and December 2020, categorized by nine unique focus areas: tracking COVID-19 risk factors and disease spread; developing disease treatment; identifying availability of and need for supplies; monitoring adherence to non-pharmaceutical interventions, or strategies to mitigate the pandemic that are non-clinical

in nature; understanding public perceptions and behavior; promoting institutional accountability and protecting human rights; addressing misinformation; alleviating pandemic-related poverty and food insecurity; and fostering trade and supporting business solvency.

The data initiatives outlined in this resource also take many different forms, using various approaches to match the supply of data and expertise with the demand for it among actors working to address COVID-19 in LMICs. One notable model involves the use of prizes, calls to action, or competition to spur distributed actors to create new and innovative uses of data to help address some dimension of the pandemic. These data challenges often involve the awarding of funding to selected applicants, but others use massive collaboration, non-remunerated competition, or clear calls to action to catalyze data innovation. Many prominent international civil society organizations, multinational corporations, and philanthropic bodies are experimenting with these new approaches and providing insights and lessons learned to help advance the field.

Since its inception, the Data4COVID19 Repository has brought together hundreds of global researchers and practitioners to crowdsource a public resource tracking the evolving use of data in pandemic response. The resource was used to inform investments and mobilization of data-driven response efforts from institutions around the world, including the New York City Government, French Development Agency, and International Organization for Migration, among others. The resource was also further developed to guide efforts in priority sectors, such as tools and insights related to Data4COVID19 efforts in the Mobility sector.

COVIDaction data challenge

In May 2020, the UK's Foreign, Commonwealth & Development Office's funded Frontier Technology Hub (Simpson, 2021) launched the COVIDaction Data Challenge (Klein, 2021), a globally inclusive call for data innovations responding to the pandemic. As a direct result of this data challenge, 12 grantees were selected to build and scale public goods, such as open data libraries, interconnectivity tools, epidemiologic models for decision-making in LMICs, and advocacy for good governance and data rights in relation to emergency data requirements. The call itself spun off several pieces of research and analysis. The challenge has also generated a wealth of data from its applicants, which provides an invaluable opportunity to explore how data needs are identified and innovations take shape across different contexts with first-hand experiences.

Aims

With limited evidence on the impact and effectiveness of data challenges to create public value in LMICs, and the rapid deployment of numerous data challenges to help address COVID-19 globally, there is a clear need for more insight into optimal design principles for data challenges that fill knowledge gaps, unlock the value of data for decision-making, and lead to impactful real-world action.

This research seeks to advance our understanding of how (1) data challenges can effectively identify and address data use gaps in the response to the COVID-19 pandemic in LMICs, (2) the data challenge model can be maximized to support innovators and create lasting impact, and (3) lessons learned from the pandemic response can inform future efforts to unlock the value of data to address emerging threats.

Methods

Three distinct sources of qualitative data are used to address this inquiry with insight from key groups in the data innovation ecosystem. These datasets include (a) a survey of innovators who design and realize solutions for data challenges, (b) a survey of the organizers of said challenges, and (c) a focus group discussion with professionals who utilize COVID-19 data for evidence-based decision-making in LMICs.

Surveys

Innovators were identified as applicants to COVIDaction's open call for innovations (Klein, 2021) addressing one of four focus areas: epidemiological systems, data collection tools, data analytics and use, and responsible data tools. In their applications, innovators submitting their work were also asked two open-ended questions:

1. *What do you see as the biggest gaps in effective COVID-19 data use?*
2. *Where do you see progress being made in COVID-19 data use?*

The GovLab's repository of data collaboratives (The GovLab, 2021) initiated in response to the COVID-19 pandemic was used to identify data challenge organizers. Researchers at The GovLab regularly conducted due diligence to identify new, innovative, and representative instances of data collaboration to address dimensions of the pandemic, and the field at large was invited to suggest new updates and additions to the repository, with curation by GovLab personnel. Representative organizers were identified from 27 data challenges in the spring of 2021 and invited to complete an online survey. These organizers were asked four open-ended questions:

1. *How did you/your organization determine target areas for the call-to-action?*
2. *How do you decide what innovations to fund or scale up?*
3. *What would you do differently if you/your organization were to put out a new call-to-action?*
4. *Has the COVID-19 data need changed since the beginning of the pandemic response (i.e., since March 2020)? If so, how has it changed? If not, how do you know the need is the same?*

Thematic analysis

All innovator responses were anonymized with an ID number by an author who did not participate in coding and a sample was then selected by coding authors for analysis. Seventeen percent (50/288) of all valid submissions were selected randomly with probability proportional to size, based on the number of submissions to COVIDaction's four key target areas (responsible data [4/10], epi models [4/17], data collection [30/178], and data analytics [12/83]). Submissions that provided complete information, were appropriate to the target area, and were not duplicates were considered valid. A sub-sample of innovator responses were randomly chosen (12% of the sample) and inductively coded line-by-line, or word-by-word (where answers were less than two sentences) for main and sub-themes by JD and AY in accordance with Braun and Clarke's (2019) framework. These codes were compiled into a single coding system, which was then applied deductively to all responses in the sample by JD and AY using MAXQDA Analytics Pro (VERBI GmbH, 2021). All identified themes, supporting ideas and examples, and their frequency of occurrence are reported for innovators in Results I.

As the initial population size was small, all responses to the organizers' survey were able to be included. The main topics discussed for each question were collated into a separate coding system and are reported in Results II.

Focus group discussion

Following the drafting of both coding systems, a virtual focus group discussion was held with members of AfriLabs, an organization with the largest and most diverse community of technology hubs, innovators, and entrepreneurs in the African innovation ecosystem with over 296 technology and innovation hubs across 49 African countries and the diaspora. AfriLabs has successfully impacted and served multiple stakeholders and clients, providing a wide range of services such as capacity building, research, networking, funding, etc., and the organization is a collaborative partner of COVIDaction. They also work with investors, banks, and other funding sources to help promote private sector responses to the needs in Africa.

Participants were recruited with the intention to represent a diversity of professions that all work with COVID-19 data in some capacity, including policymakers, government officials, entrepreneurs, and industry experts. As Kenya, Nigeria, Uganda, and Ghana were the most common African countries of origin for submissions to COVIDaction's Data Challenge (22, 18, 11, and 9, respectively), AfriLabs members working in any of these countries were invited. There was no overlap between innovators, organizers, and discussion participants.

The aims of this event were to discuss the capacity for data challenges to identify and address gaps in data use for informed decision-making, based on participants' expertise. In small break-out rooms, participants reviewed preliminary findings (the complete coding systems) and described themes or topics that were missing; details or examples to add; and organizational/hierarchical changes to ensure the codes were relevant and inclusive. In the following main session, all participants identified strategies for translating data insight to impact and how data challenges can support this process. Key topics raised during this event are reported in Results III.

Results

The results of the thematic analysis of the innovator survey are presented first, divided into Gaps and Progress. Then, topics addressed in the organizers' survey are presented, followed by the findings of the focus group discussion. Each set of results are presented without significant author interpretation beyond that required for a cohesive write-up. Explanation, context, and examples provided are from COVIDaction applicants (innovators), survey respondents (organizers), and focus group participants (professionals), respectively.

Thematic analysis: Gaps

Eight main themes were identified among innovators' responses to COVID-19 data gaps. Each main theme is listed with its sub-themes (where applicable), and the number of responses in which a theme occurred at least once is given in parentheses. All main and sub-themes required mention five times or more for inclusion. Quotes representing the main theme by elaborating on sub-themes and/or providing additional nuance are also included.

Data quality and accessibility limitations (44) occurred the most frequently, in 88% (44/50) of all responses. The most common sub-themes of these limitations were data quality¹ (19); the availability of real-time or timely data (16); publicly accessible or open-source data (18); follow-up data (9); and under-utilized sources (8) such as social media or digital trace data. The lack of centralized information management systems (13) was also mentioned by 26% of innovators, which indicates how these data limitations can go unaddressed and why they are so challenging to resolve at a regional or national level. In the absence of a centralized system, Respondent 1126 describes what precisely the public are missing and the adverse effects of reliance on social media as the most accessible information platforms:

There is also a lack of platform where patients can submit their experiences of symptoms and get timely advice on which decision to make unless they visit a health care facility but also the overall information management of lockdown times and processes, precautions to take and other vital information is transmitted through social media which creates controversies and several claims of cures which may encourage self-medication and lead neglect of established primary care institutions and professionals. (1126)

Human resources and institutional capacity (32) were common barriers to effective data use in LMICs. Specifically, this limited capacity was due to slow, burdensome data entry (10); not enough frontline

¹ The authors define data quality as data (qualitative and quantitative) that meets programmatic needs for "validity, integrity, precision, reliability and timeliness" (USAID Conducting Data Quality Assessments). It is understood that different decisions have different data quality needs; however, across the board, interviewees noted that they lacked data of sufficient quality to meet their programmatic needs.

health workers (9); limited data literacy among policy/decision-makers and the general public; and an absent or outdated understanding of data protection needs (6). Eight respondents specifically mentioned paper-based data collection and manual entry as bottlenecks in its processing, analysis, and reporting. This finding demonstrates how data may go unused or quickly become outdated. Respondent 1165 provides additional detail on this data need:

Even before COVID-19, health and human services have lacked real time visibility into their actions and impact - often capturing data on pen and paper or through cumbersome data entry tools. Due in part of the lack of capacity building digital grants or investments into those infrastructures. Now, as COVID-19 forces digital transformation in every sector, LMICs need to advance their support systems and digital service delivery more than ever. (1165)

Evidence-based decision-making (21) is not possible without reliable and updated information. Aforementioned data limitations and bottlenecks impact the ability of policymakers in LMICs to effectively plan interventions and allocate resources. Data are also limited in their capacity to allow public health researchers and decision-makers to conduct ‘deep dives’ investigating COVID’s effects on other health and social topics in a particular context. This limitation further inhibits a government’s capacity to address the immediate and long-term impacts of the pandemic. Respondent 1156 notes that this limited information will bottleneck ‘political will’ of governments during the pandemic:

With limited real-time information on COVID-19 cases, attributed deaths, and relevant awareness levels and behaviors, it’s difficult to have a ‘ground truth’ to get a grip on the epidemic. We see amazing political will across Africa, but it’s hard for countries to know where they should be allocating their limited resources. (1156)

Impacts on vulnerable communities (20) were specifically cited by 40% of innovators, the most common of which was rural residents (8). Remote individuals are poised to suffer greater impacts of mortality and economic hardship overall, as healthcare and resources are more difficult to access, particularly with pandemic-related supply-chain interruptions. Respondent 1171 notes that remote locales are a significant barrier for initial risk management and disease surveillance:

The turnaround time of testing for rural settings is much worse than average since most of the testing equipments are based in major government facilities. With a majority of the population living in rural settings, this delay puts the majority of the general population at risk of contracting COVID-19. (1171)

To address some limitations of surveillance data from rural settings, Respondent 1066 recommends decentralizing the approach through:

...data collection, analysis, decision-making, and data sharing at the village and health center level where the majority of covid-19 cases will likely present once there is community spread of the virus. (1066)

COVID-19’s exacerbation of existing difficulties in LMICs (16) also poses a substantial concern for respondents. Particular strain is put on limited or outdated digital infrastructure and policies (14); connectivity and internet access (10); poverty (9); the neglect or worsening of national situations regarding other diseases (6); and malnutrition (5), often as a result of interrupted supply chains. Respondent 1138 notes how smaller-scale entrepreneurs and local businesses in LMICs are now at a greater disadvantage using the example of how global logistical support systems were not set up to meet the needs of LMICs, even before the pandemic:

Aid Supply chains and logistics are often EU- or US- centric. With current collapsing supply chains due to Covid-19 this heavily affects LMICs; aid agencies and local governments. Local

businessmen are even more affected since backup financial support as is available for western countries often are not available in LMICs. (1138)

A lack of institutional coordination and technical integration (16) impacts the interoperability of tools that collect and process data (6); inhibits collaboration amongst multidisciplinary stakeholders (5); and slows knowledge-sharing and collaborative action across labs and clinics (5). Effective tools are those that align well with overarching needs for a specific context and do not silo the data they gather or insight they generate. Sustainable tools will also adapt to serve the population after the event. Respondent 1081 provides further detail:

LMICs are more reluctant now than ever to add new data systems to their health information system landscape unless they have the ability to be integrated or be made interoperable with existing systems. They are insisting on donor and partner coordination in resource allocation and program implementation, reducing redundancies in the number and types of digital solutions and technologies welcomed. Finally, especially in the context of disease outbreaks and other humanitarian crises, LMICs leaders increasingly require that digital solutions be sustainable and locally owned. (1081)

Lack of predictive data (12) was commonly noted as a substantial gap, inhibiting stakeholders' ability to take a proactive rather than reactive posture in their response; preventing transmission and supporting vulnerable populations before situations become more precarious is the most effective option to save lives. However, this option does not exist without data of sufficient quality to allow for reliable prediction and planning. In overcoming this gap, Respondent 1165 cautions:

Until there is a transition from data collection to digitally enabled program delivery, health and human service providers will be looking in the rear-view mirror. COVID-19 makes it more clear than ever that rear-view mirror optics are not adequate to support organizations to deliver the right care or service at the right time. (1165)

Insufficient data protection (11) emerged as a main theme that closely ties with each of the others. Personal data privacy and security processes and procedures (8) and formal data protection regimes (5) were identified as important gaps in LMIC data management ecosystems. Many data protection regimes are weak, outdated, or absent, which increases the risk of mismanagement. Yet the need for sensitive data to mitigate the pandemic, including health and movement information, is increasingly glaring. Respondent 1054 summarizes the need on all sides:

The current pandemic has brought to light the need to develop data sharing frameworks to help address COVID-19 and future public health outbreaks, while also making sure the privacy and human rights of citizens are protected. (1054)

Thematic analysis: Progress

Seven main themes spanned innovators' responses to areas of progress in COVID-19 data use. Many were supported with specific examples that did not translate into common sub-themes and necessitated anonymization, and therefore fewer sub-themes are reported than in Results I.

New fit-for-purpose (FFP) tools (28) are a major area of progress identified by 56% of innovators, which can include applications for COVID-19 data collection, public outreach and education, and epidemiological and statistical models. The value of bottom-up approaches in designing FFP tools is described by Respondent 1069:

Significant progress has been made by organizations understanding that a bottom-up approach to data collection is key to getting the best results. Introducing new digital solutions requires patience when training end users on new technology, but most importantly it requires involving end users

feedback in the design and configuration of data collection tools to make sure end users understand how important their work is to achieve organizational wide KPIs. (1069)

Another benefit has been that data quality expectations and requirements are designed into the tool, with “at collection” data validation opportunities. This area also benefits as tools are tested and validated around the world. New tools are developed globally to align with modern standards of quality, accountability, transparency, and inclusion. The holistic benefits of this trend, including the value of bottom-up approaches and the need for KPIs, are highlighted by Respondent 1069:

Significant progress has been made by organizations understanding that a bottom-up approach to data collection is key to getting the best results. Introducing new digital solutions requires patience when training end users on new technology, but most importantly it requires involving end users feedback in the design and configuration of data collection tools to make sure end users understand how important their work is to achieve organizational wide KPIs. (1069)

An important caveat is brought up by Respondent 1066, who elaborates on the importance of prioritizing adaptable, sustainable tools, but cautions on potential negative impacts of overly specific, targeted approaches:

Tools created to address one specific emergency such as Covid-19 often exacerbate the situation by too narrowly focusing on the current crisis without considering the entire ecosystem within which this tool will function. This leads to a tool that may be highly effective for the specific need, but which exacerbates silos within systems, impedes the delivery of routine care services and becomes irrelevant as soon as the emergency has passed. (1066)

Expanding inter-organizational and inter-sectoral partnerships (18) was a benefit associated with increased virtual socialization and a recognized need for information sharing, both internationally (10) and across disciplinary silos (6). The benefits of a higher demand of shared knowledge are felt across sectors. The resulting collaborations accelerate innovation and uniquely provide for different needs, as noted by Respondent 1156:

1156: As LMICs were slower to get hit by the pandemic, insights obtained from the earliest-hit countries - such as epidemiological risk factors and the effectiveness of early lockdowns - can be used to guide policy in low-resource settings.

COVID-19 highlighted data needs (16) and encouraged broader discussion around data collection and use, with space to plan for the potential impact on an individual’s privacy and human rights. The pandemic has underscored increasingly pertinent needs for data protection regimes and systems, and as Respondent 1036 also notes, opportunities:

The COVID-19 pandemic is an opportunity from a responsible data viewpoint to do two things. Firstly, to highlight the power (both positive and negative) of modern data-driven technologies. Secondly, the crisis has highlighted the extent to which data-driven technology can be used to enforce surveillance measures and curtail fundamental rights. This in turn contributes to healthy debates around the limitations of such approaches. In countries with non-existent or nascent data protection regimes, these debates can help to frame their future development and also help to foster an expectation of accountability. (1036)

The establishment of public health/research infrastructure (13) in LMICs has widely supported data quality and availability. This theme includes the prioritization of surveillance data collection (19); data sharing and reporting (17); monitoring evaluation (10); communication to public/media (10); testing/screening and triage (5); and national public health institutes (5). Often these systems had been bolstered during the Ebola epidemic, and lessons learned in that context could be applied to support the COVID-19

response. In line with respondent 1066's quote on surveillance, respondent 1150 notes how some systems already integrated with existing public health infrastructure can be maximized further:

Established existing health provision channels, health record systems, mobile networks, geospatial, and workplace employee monitoring systems might be leveraged to collect the necessary and reliable information for workers' health insights. (1150)

Clear value proposition of spatial data (10) collection and use was noted by 20% of respondents as an area of progress that supported awareness of geographic trends in the disease's spread, as well as subsequent health and social effects. Increasing the availability of spatial data is important for making data interoperable. Novel uses and sources for spatial data have proven valuable in tracking, forecasting, and preparing, as identified by Respondent 1048:

As societies across the world are discussing potential exit strategies from covid-lockdowns, governments are increasingly considering how to use data to track movement and behaviors of citizens in order to prevent further spread of the virus. (1048)

An increasing understanding of data sensitivity (6) among governments and citizens has shaped the market to favor tools that use data more responsibly. Though the progress of these trends varies by context, overarching governance norms regarding data use, highlighted by Respondent 1036, must be upheld by tools that aim for adaptability:

...although a plethora of tools exists, context is crucial. Prerequisites to the successful deployment of any responsible data policy tool will therefore be built-in adaptability and flexibility, an understanding of the need for multi-stakeholder partnerships in creating governance norms (including in the field of data governance) and an appreciation that transparency, inclusion and accountability must be integral to data use as part of the response. (1036)

Education of the general public (6) has combated misinformation and supported the effectiveness of non-pharmaceutical interventions throughout the pandemic. In many cases, consistent public education and communication remains an obstacle but improvement in this area is prioritized by LMICs as a crucial element of their pandemic response. Respondent 1045 explains the value of this:

Effective communication with the public is an important factor for epidemic control and is key to managing public health emergencies in order to inform citizens, to share information and to provide guidance on risk and exposure mitigation. (1045)

Role of data challenges

Seven organizers of COVID-19 data challenges provided insight into the process of managing an open innovation call and the role of supporting ground-up solutions in the data innovation ecosystem.

In determining target areas for their calls to action, organizers report speaking with interdisciplinary partners and engaging with stakeholders, including government officials, senior data scientists, and health experts. Focus areas were also determined based on data needs identified by partner institutes. In two cases, organizers kept their call very open and advertised the focus area as simply the intersection of broader topics in relation to COVID-19, like technology and policy, or the social impacts of the pandemic.

Multiple strategies were used to allocate funding, resources, or expertise to scale innovations. Organizers commonly employed review panels comprising domain experts and local stakeholders. Innovations were also judged on the clarity of their pathway from insight to action and the level of expertise among the team; organizers often required detailed documentation or proof of concepts. The urgency and priority for each innovation was also considered.

Though organizers expressed happiness with the innovations catalyzed by their challenges, they identified some changes to be made in future calls to action. Several noted that innovators should be asked to provide greater detail in their initial proposals or applications. Similarly, organizers believed that they should provide prospective applicants with a more specific framing of the challenge; a clearer understanding of the level of time needed; and targeted metrics of success. Organizers were interested in providing more funding; more support for early-career researchers; and space for ongoing partnerships and coordination across groups with similar work areas to resist information siloing. While wishing to do more, it was also noted that more volunteers would be needed on the organizers' end. An overall shift in focus was mentioned by one organizer, with a view to how changes in the pandemic would affect a future call-to-action:

Going forward, we would advocate for more effective data challenges aimed at increasing preparation, readiness, and prediction of future needs and opportunities in the interest of helping society become better able to proactively address emerging threats rather than finding ourselves in a reactive posture. (44)

Most organizers found the COVID-19 data need has changed since the beginning of the pandemic response. For example, expectations of data quality have risen (and in some cases, data quality has also improved). New datasets have also been created by the emergence of variants and the developing understanding of the virus; the integration of vaccines; non-pharmaceutical interventions deployed for different time frames in different contexts; and the ability to study longer-term effects of the pandemic. These new factors have not replaced initial data needs, but rather compounded the overall need for insights to inform timely public health decisions.

Insight to impact

Participants in the focus group discussion identified three key topic areas that must be addressed to translate data insight to impact in the pandemic. These topics are supported by explanations and examples provided here by discussion participants.

Reaching vulnerable populations and remote areas remains a significant difficulty. Governments and INGOs need to engage more directly with the rural community and religious leaders around addressing COVID-19. Building capacity of external or national actors to understand the particulars of local communities will address the lack of partnerships and work to de-silo information. Further, community leaders can help make the case for localizing the promotion of data generation and collection. For many communities, urgent sectors are being overlooked because of the need to react quickly. The scenario of high-income countries reacting to COVID-19 has shifted focus away from addressing other communicable diseases that remain highly prevalent and place vulnerable populations in yet a more precarious position. A unique opportunity was also identified in this case, which is the importance of historical data use cases. Participants cited how research around malaria transmission and risk mapping in Nigeria has provided data essential to proactively identify communities with higher risks of morbidity and mortality for COVID-19. Lessons learned from malaria research further support effective monitoring of COVID-19 and mitigation efforts in vulnerable populations.

The public's education is a critical factor for any innovation or intervention to be effective. The importance of data collection is not clearly understood by the public and there is often mistrust about the intended use of the data. In Uganda, the data management value chain is lacking because of mistrust of data sharing and existing policies. As a result, much useful data are not collected or go unused, and the data that are collected are not always of high quality. This barrier also presents an opportunity to improve data quality through localization of collection and communication. For example, Public Health Centers are the first point of contact for most in Nigeria, but these centers are not being used for testing, resulting in significant data gaps and insufficient reporting. Increasing capacity for local, trusted health resources to address COVID-19 needs in the community would also enable more individuals to seek testing and care

when needed. Supporting these centers to play a bigger role in mitigating the pandemic could also slow the spread of misinformation. At present, social media has skewed perception of pharmaceutical interventions; where local health centers are seen as trusted authorities on COVID-19 information, and clear paths of communication with the centers are well-established, community members will be better informed and know where to seek proper medical advice. Limited public education cuts across obstacles of delivering impact due to false information and poses a complex problem for policymakers who need the public's trust and participation to carry out effective interventions. In all of these scenarios, participants further stressed that education is key to obtaining complete, accurate data, and supporting the integration of solutions addressing the pandemic.

Innovation ecosystems in LMICs have space to be maximized in the wake of the pandemic. COVID-19 has presented a clear opportunity and value proposition for an increased focus on and investment in social innovations in health provision. Simple, applicable solutions are being reinforced by public use, which has given governments motivation to work on solutions that would not otherwise have been prioritized and (in learning from innovators and global examples) the confidence it can be done. Innovators will therefore benefit from working around a national innovation-developing cluster program. By letting the government take center stage, integrating and scaling innovations become easier due to government involvement. However, ethical assessment frameworks are still necessary to determine the level of access to data that governments should have, as this access can carry unintended consequences that need to be addressed. Participants noted that, in LMICs, and Sub-Saharan Africa in particular, scalability is difficult because innovations must address priority areas set by the government. Innovation ecosystems have seen the emergence of new enterprises throughout the pandemic as projects adapt to new issues and align to new norms. Businesses have diversified and adapted to remote work; this switch to virtual has meant more global participation, enabled a wider reach, and lessened the cost of doing business in many cases. Many existing solutions were repurposed (like ventilators), or innovators did a U-turn to produce something else that was needed. However, to enable more innovations to scale, innovators must understand how and where to connect to get resources; connectivity itself remains a huge problem and information networks must also be readily available. Adherence to standards needs to be maintained, regulatory approvals obtained, and global supply chain barriers must be overcome. Countries with more established innovation ecosystems can fast-track this work, but for many innovators in LMICs, it is difficult to know where to start and how to compete at the global stage. Furthermore, top innovators are utilizers of strong intellectual property systems; there are issues of open data and open science for innovators budgeting what can be kept proprietary and what can be given away. Open innovation has been critical in the initial response to COVID-19, but the funder, rather than the innovator, often claims ownership and, in some cases, intellectual property rights over the data generated. Producing products locally; connecting innovators with global networks and resources; streamlining the innovation scaling pipeline; and developing considerate intellectual property regimes are important elements required to close this gap in LMICs.

Discussion

Many of the gaps identified by call-to-action respondents, organizers, and discussion members are applicable beyond the immediate context of the COVID-19 pandemic and highlight the importance of a healthy data ecosystem more generally. To promote impactful and responsible data use, a data innovation ecosystem requires the human and institutional capacity necessary to use data effectively and protect potentially sensitive information streams. Institutional data literacy and competence represent core areas of need to help unlock the potential of data (and data challenges) in the response to COVID-19 and in the broader effort to leverage data to create new public value and address public problems. Yet human resources and institutional capacity can only go so far in the absence of timely, high-quality data and FFP, integrated data collection, analysis, and reporting systems.

A common denominator in these data quality and institutional capacity issues lies in the availability of funding. The limitations arising from the lack of funding were evidenced throughout our results both implicitly and explicitly. Respondents made clear that a well-developed digital infrastructure connecting

key actors is essential to ensure data are actually reaching people who are capable of using it in meaningful, responsible ways. Respondents noted that without good data and communication, social media tends to fill the gap, allowing misinformation to go unaddressed, spurning misinformed decision-making, and ceding control over the situation to misinformed and/or malicious actors. Clear legal and policy guardrails that provide a sense of what can be done, and how, are crucial guides for supporting a healthy data innovation ecosystem. After proving effective, these guardrails may be more internationally relevant as a framework than a common system of measurements to establish ground truths, given national sovereignty and varying regional contexts/needs. Overall, these guides are essential to increase impactful data use and reduce the risk of data misuse. Good data protection may incorporate data minimization and anonymization, access controls, and ethical assessments by independent bodies to maintain fairness and security. Clear and effective data protection laws and policies are needed to help guide responsible data handling, but these policies should also be designed and implemented in a way that advances innovative and impactful data use and reuse, rather than inhibiting the potential of data to provide transformative impact in the response to pandemics and other emerging public crises. In supporting individual literacy to navigate complex systems, encouraging communication between stakeholders, prioritizing systems integration, and developing comprehensive, context-specific policy guardrails, our findings demonstrate how good data management principles align with good public health practice.

The role of FFP tools in the pandemic was highlighted by all groups in our inquiry, with frequent, specific mention of sustainability through adaptability. FFP tools can impede progress by being too narrowly focused and contributing to ongoing obstacles of siloed information. Yet they can strengthen existing systems by addressing current emergency needs if they can also be useful in the long term or adapted to different contexts. This aspect of sustainability is enabled by expanding partnerships, allowing more data to be used to refine a tool, and ensuring tools are foundational to future data use (e.g., using spatial data). Though definitions of sustainability vary, there is consensus that a tool's future utility and return on investment are affected by integration and consistent funding. Future challenge organizers, funders, and innovators will be well-served in considering how to strike the right balance between hyper-customized tools for the issue at hand, and more general-purpose and broadly applicable data systems and innovations. User-driven innovation.

Across all groups, there was significant discussion around sustainable public good tools (such as open-source tools and open data investments) and tech. Participants, organizers, and professionals increasingly recognize these tools need to be more than the sum of their parts, contrary to preconceived notions about what data challenges typically do not do. The need to avoid a common pitfall in an unhealthy data innovation ecosystem, namely the re-siloing of information within each individual project, was described across all groups. During the pandemic, the industry has been increasingly incentivized to build open-source tools for public good, which is creating a foundation upon which others can draw. Indeed, key data gaps such as a lack of predictive data and insufficient data protection are also extremely relevant in high-income settings. The next steps to progress healthy data innovation ecosystems will involve clarifying and codifying an enabling intellectual property environment, streamlining innovation scaling processes, and setting up innovation target areas with government involvement.

COVIDaction's work and research also demonstrate how data challenges can be effective in setting the agenda by clarifying important considerations, mobilizing work, and building awareness across key stakeholders and actors in the data ecosystem by elaborating priorities, evaluation factors, and definitions of impact. They are useful in motivating actors in different contexts to take on work aligned with the call-to-action. To reflect this aim and recognize that the connections these programs can build are often as important as funding, COVIDaction uses a partnership-driven model; in the experience of the authors, this framework is the sustainable choice to produce long-lasting impact. The nine challenge areas of our call-to-action also overlapped with existing tensions in this innovation space: the clear need for innovative, predictive work around data, the need to have a clear sense of supply to existing data, strong connections to stakeholders, and connections to the broader ecosystem and infrastructure. This approach also allowed

us to support innovations that had great efficacy to address their most crucial concerns while contributing overall to a healthier data ecosystem.

Limitations

Our qualitative data has benefitted from the input of key actors at multiple levels, each with open-ended opportunities to share what problems are most crucial for them, or what solutions are most practical. However, our datasets are not without limitations.

The representativeness of each included group should be explored. The initial call-to-action sought mature applicants, and therefore was limited to those whose work was underway. The survey for call-to-action organizers was administered to all qualifying representatives, but this group is already a small population. The survey also had a low response rate (about one-third), and so a standard thematic analysis was not possible. The FGD invitations were limited to four countries in Africa with the intent to facilitate the discussion through more similar practical contexts, which were also shared by a substantial portion of our applicants. However, applicants from other LMIC regions may not have had their contexts fully represented in the discussion. And while gender diversity was evenly reflected in our invitations, few women from AfriLabs were able to attend the event due to scheduling conflicts.

The structure of the call-to-action from which the applicant responses are drawn must also be considered. For applicants, answers to the gaps and progress questions were limited to 2000 characters and their answers may tend to be self-serving so their applications can look good; gaps discussed tend to be in terms of what applicants think they can solve, so there is a natural link between gaps and progress as reported here. Regardless, the gaps raised are valid and relevant to the specified context. There was less discussion of the need for data to address economic issues or broader gaps in the space, which may be due to the call's focus on tools immediately benefiting individuals. The variation between groups of innovators responding to a particular call should also be noted; data protection came out more strongly in our challenge because responsible data items were explicitly requested. However, any innovators seeking to create a healthy data ecosystem would see these items addressed throughout other types of investments.

Finally, all research (submission, survey, and focus groups) was conducted in English which may have excluded respondents initially or limited how they were able to express themselves.

Conclusion & Recommendations

For funders and policymakers

Funders and policymakers can demonstrate commitment to innovators and responsible data use by supporting the integration of the innovation with existing physical, digital, and policy infrastructure. Supporting bottom-up approaches that follow a partnership-oriented model can also engage with early-career innovators by connecting them with funding, entrepreneurial advice, and multidisciplinary professional networks. These approaches should be supported with unified principles of good data and digital governance, including a focus improving data quality and access while also protecting this data against misuse. This further enables the strong integration of the innovation with the targeted context and subsequently its efficacy.

For data holders

Timely, available data that does not cause harm is vital to a healthy data innovation ecosystem. Data holders can enable innovative, responsible data use in their sector by making their data as accessible as possible. Data holders are also responsible for abiding by well-developed data protection rules and principles and training individuals to manage data securely and fairly, balancing this against the need for more precise, granular data which is foundational for data quality improvement.

For innovators

Overall, sustainable, successful tools are based on principles of good governance, flexible enough to be useful in different circumstances or crises, and effectively integrated with the needs and capacities of their target setting. For example, successful innovations often improve access to quality data that can improve decision-making by key actors. Securing initial funding and prioritization in the data innovation ecosystem is supported by demonstrating clear and appropriate strategies around decision-making and an understanding of how to use data responsibly, who will use the innovation, and how.

Acknowledgments. Authors are grateful for the help of many colleagues, including Patrick Ashu, Sarah Goodwin, and Alexis de Bruhoven for additional project management and focus group discussion organizational support. We are also especially grateful to the members who participated in our focus group discussion and shared their time, expertise, and insights, several of whom wished to be identifiable here: Henry Agyei Asare, Tentmaker Hub, Ghana, henry@tentmakergh.com; JohnBosco Ugochukwu Ezenwa, Emerge-Data Integrated Services Ltd., Nigeria, ezenwaugoj@gmail.com/ezenwa@ediservices.com.ng; Patrick Joram Mugisha (Moogy), Ministry of Science, Technology and Innovation, Innovent Labs Africa, Uganda, patrick.mugisha@innoventlabsafrica.com; Gilbert Junior Buregyeya, Makerere Innovation and Incubation Centre (MIIC), Startup Uganda, Uganda, juniorgilbert18@gmail.com/gburegyeya@miichub.com.

Funding Statement. Research design and manuscript writing was provided by Jamie Danemayer on behalf of the Global Disability Innovation Hub in her role as Researcher, supported by the AT2030 Programme which is funded by UKAid, project number: 300815 (previously 201,879–108). However, no specific funding for this study exists. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests. The authors declare no competing interests exist.

Author Contributions. Conceptualization: J.D., A.Y., and M.K.; Data curation: J.D., A.Y., S.G., and M.K.; Formal analysis: J.D. and A.Y.; Funding acquisition: A.Y., S.G., and M.K. Investigation: J.D., A.Y., S.G., and M.K.; Methodology: J.D., A.Y., and M.K.; Project administration: A.Y., S.G., and M.K.; Resources: A.Y., S.G., L.E., and M.K.; Software: J.D. and A.Y.; Supervision: M.K.; Validation: J.D. and A.Y.; Visualization: N/A; Writing—original draft preparation: J.D. and A.Y.; Writing—review & editing: J.D., A.Y., S.G., L.E., and M.K. The sponsors had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement. Restrictions apply to the availability of some of these data. Data obtained from COVIDaction (innovator survey) and generated by this study (organizers survey and focus group discussion) are available from the corresponding author with the permission of COVIDaction, due to data protection agreements between respondents and researchers. Data obtained from the GovLabs repository are publicly available from list.data4covid19.org.

Informed Consent. Informed consent was obtained from all subjects involved in the study.

Institutional Review Board. This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the University College London Research Ethics Committee (approval number 1106/014, 13/08/2020).

References

- Ågren K, Bjelkmar P and Allison E (2021) The use of anonymized and aggregated telecom mobility data by a public health agency during the COVID-19 pandemic: Learnings from both the operator and agency perspective. *Data & Policy* 3, e17. <https://doi.org/10.1017/dap.2021.11>.
- Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, Seneviratne A, Hu W, Janicke H, and Jha SK (2020) A survey of COVID-19 contact tracing apps. *IEEE Access* 8, 134577–134601. <https://doi.org/10.1109/access.2020.3010226>
- Angarita MAM, Nolte A (2020) What do we know about Hackathon outcomes and how to support them? – A systematic literature review. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Tartu, Estonia: Collaboration Technologies and Social Computing: 26th International Conference, CollabTech 2020, 12324 LNCS:50–64. https://doi.org/10.1007/978-3-030-58157-2_4
- Braun V and Clarke V (2019) Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*. 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- The GovLab (2021) Data4COVID19: A Living Repository for Data Collaboratives Seeking to Address the Spread of COVID-19. Available at <https://list.data4covid19.org/> (accessed 20 July 20 2021).
- Gama K (2021) Successful models of hackathons and innovation contests to crowdsource rapid responses to COVID-19. *Digital Government: Research and Practice*. 2(2), 1–7. <https://doi.org/10.1145/3431806>
- Harrison TM and Pardo TA (2021) Data, politics and public health: COVID-19 data-driven decision making in public discourse. *Digital Government: Research and Practice* 2(1), 1–8. <https://doi.org/10.1145/3428123>

- Jagadish HV, Stoyanovich J and Howe B** (2021) COVID-19 brings data equity challenges to the fore. *Digital Government: Research and Practice*. 2(2), 1–7. <https://doi.org/10.1145/3440889>
- Kishore N** (2021) Mobility data as a proxy for epidemic measures. *Nature Computational Science* 1(9), 567–568. <https://doi.org/10.1038/s43588-021-00127-7>
- Klein M** (2021) Building a Healthy Data Ecosystem to Fight COVID-19 #DoingDataRight. Available at <https://medium.com/covidaction/building-a-healthy-data-ecosystem-to-fight-covid-19-or-doingdataright-ed0305bd1313> (accessed 20 July 2021).
- OECD** (2020) OECD Policy Responses to Coronavirus (COVID-19): Protecting People and Societies. Available at <https://www.oecd.org/coronavirus/policy-responses/covid-19-protecting-people-and-societies-e5c9de1a/> (accessed 20 July 2021).
- Redmiles EM** (2021) User concerns 8 tradeoffs in technology-facilitated COVID-19 response. *Digital Government: Research and Practice*. 2(1), 1–12. <https://doi.org/10.1145/3428093>
- Röddiger T, Beigl M, Dörner D and Responsible BM** (2021) Automated data gathering for timely citizen dashboard provision during a global pandemic (COVID-19). *Digital Government: Research and Practice*. 2(1), 1–9. <https://doi.org/10.1145/3428471>
- Simpson L** 2021 About Frontier Technologies Hub – Medium. Available at <https://medium.com/frontier-technologies-hub/about> (accessed 20 July 2021).
- VERBI GmbH** (2021) MAXQDA Analytics Pro 2020 Student.