



The Use of **Human Mobility Data** in Public Health Emergencies

Satchit Balsari, Caroline Buckee, Jennifer Chan,
Andrew Schroeder

About

CrisisReady

CrisisReady is a research response platform based at Harvard University and Direct Relief that collaborates with academic institutions, technology companies, and response agencies to embed data-driven decision making into local disaster planning. CrisisReady develops scalable expertise in Data Readiness, Methods Readiness, and Translational Readiness by securing data pipelines that provide actionable analyses to meet the pre-articulated needs of policymakers and response agencies.

Acknowledgements

The organizers of the “Safe, Responsible, Fair and Equitable use of Human Mobility Data” seminars and the authors of this white paper are grateful for the support and guidance of their colleagues and collaborators. Without the support of The Radcliffe Institute for Advanced Study at Harvard University, the Global Facility for Disaster Reduction and Recovery (GFDRR), Direct Relief, Harvard Data Science Initiative (HDSI), the Center for Communicable Disease Dynamics (CCDD), and Beth Israel Deaconess Medical Center, as well as the rapporteurs who participated in the seminars, this work would not have been possible. The views expressed in this report, the underlying analysis, and any remaining errors are the authors’ alone.

Table of Contents

Executive Summary	3
Data Readiness	6
Provision of Access	6
The Mechanics of Access	7
Resolution vs. Utility	10
The Saliency of Context	10
Consent and Its Limitations	11
Negotiating Access	12
Action Points	13
Methods Readiness	14
Understanding Bias and Estimating Representativeness	14
Quantifying Harm	15
Standardization	16
Combining Mobility Data with Epidemiological Data	16
Action Points	17
Translational Readiness	17
Supply-Side Expansion	18
Demand-Side Expansion	19
Action Points	20
Conclusion	22
References	23
Appendices	27
Action Points: Overview	27
Seminar Goals	28
Seminar Participants	29
Organizing Team	30





Executive Summary

This white paper summarizes key learnings from a four-part seminar, “Safe, Fair, Equitable and Responsible Use of Human Mobility Data,” which convened in March and April 2021 at the Radcliffe Institute for Advanced Studies at Harvard University. The seminar was attended by over 40 domain experts representing academia, industry, law, humanitarian relief, and disaster response. The interdisciplinary exchange sought to map areas of convergence between technology companies that produce human mobility data, epidemiologists and public health practitioners that incorporate these novel data streams into their models and research, lawyers, ethicists, and data scientists that are concerned with responsible data management, including privacy protection for individuals and groups, and health agencies and disaster responders that use insights from such data for decision making. Seminar deliberations also identified the technical, regulatory, and translational gaps that preclude the effective integration of mobility data into field response.

The advent of mobile phones and internet-connected devices has generated enormous amounts of data on individual and group mobility patterns, collected by telecommunications companies, smartphone apps, data aggregators, and brokers. For the past several years, these data have helped researchers estimate population movement patterns to inform epidemiological modeling, situational awareness, and resource allocation in crisis settings.²⁻⁶ Though these data are routinely collected by telecom and other companies for business analytics, they are shared with researchers or policymakers on an ad-hoc and limited basis. Strict national and regional legal frameworks guide the re-use of these data globally, and when the data are shared with researchers, they are done so in accordance with local law and after prolonged contractual negotiations.⁷ Data use agreements take a long time to formulate due to the unfamiliarity of regulatory bodies, ethics review boards, and data providers with the applications of these novel data streams to public health issues and the related risks and necessary protections associated with these data

sets. During a public health emergency, small pools of academic researchers and policymakers have access to these data through preexisting relationships with technology companies.⁸ With the development of “differential privacy” technology for producing aggregated data with strong de-identification, companies have begun sharing their datasets with a wider community of researchers; and some data are now available publicly.⁹⁻¹²

Access to human mobility data increased exponentially in 2020 during the COVID-19 pandemic, when non-pharmaceutical interventions like travel bans and stay-at-home orders became the mainstay of public health response. Researchers around the world used location or movement information derived from telecommunication data, such as Call Data Records (CDRs) or x-Data Records (xDRs), the latter generated with a mobile device connected to the internet; first- or third-party Software Development Kit data collated from smartphone apps; vehicle GPS devices, Bluetooth exchanges or geotagged social media data, to study the impact of non-pharmaceutical interventions on population movement and on the evolving circumstances of the pandemic.¹ Publicly accessible “scorecards” attempted to rank counties and neighborhoods based on their mobility patterns.¹³

Given the urgency of the moment, the efforts to leverage these data were laudable but fraught with limitations and potential for inadvertent harm. The provenance of these data was poorly understood by many who sought to apply the data to inform public policy. The owners of mobile devices who generate data sets are not usually a random or representative sample of the population of interest, and not all data providers have the same spatial or temporal coverage.¹⁶ Information on the representativeness and coverage of these data is rarely available and must be inferred by researchers themselves. Additionally, the methods to collect, de-identify, and share data vary widely across companies. Robust analysis needs to consider the uncertainty and bias associated with these data.

The technical expertise to conduct such an analysis is often inaccessible to policymakers and is only available to researchers that have previously worked with mobility data.

The capacity, expertise, and mandate of these researchers is highly variable, poorly mapped, and is being relied on in the absence of common reporting and recording requirements by regulatory agencies. Even if the analysis produced was robust, there remains a wide translational gap between the complex methodological questions that interest researchers and the simple, actionable information that policymakers need in times of crisis. Information generated from these data is often presented in ways that do not align with existing ways of working within the domain of emergency response. Under-sourced public health and disaster response agencies often do not have the necessary internal capacity to engage with a complex analysis during a crisis.

The “Safe, Fair, Equitable and Responsible Use of Human Mobility Data” seminars sought to identify the aforementioned challenges in accessing, understanding, analyzing, and applying human mobility data before and during emergency events to develop a shared roadmap of priorities for scientists, policymakers, and technology companies. The technical, regulatory, and societal challenges explored during the seminar and in this white paper are organized around three clusters: Data Readiness, Methods Readiness, and Translational Readiness.

The seminar series resulted in this white paper, which summarizes key points of consensus and recommendations. This document will in turn set the agenda for a consultative process (in collaboration with CrisisReady and the Global Facility for Disaster Reduction and Recovery, GFDRR) to develop guidance for governments and response agencies seeking to use these novel data streams for emergency preparedness and response.

The section on Data Readiness examines the technical, regulatory, and ethical issues concerning access to human mobility data generated from mobile phones. Key themes that emerged include the criteria and eligibility for access, the granularity of the data that can be shared, the tension between risk and utility of the data shared, the means for and barriers to sharing data across institutions or jurisdictions, and finally, the determination of the arbiter of these decisions.

The section on Methods Readiness examines issues of

representativeness, uncertainty, privacy, and epidemiological applications of these data. This section outlines advances in the application of multiple large data streams generated by mobile digital advertising (AdTech) companies, social media platforms, and telecom companies to public health response planning and modeling. We list potential technical and regulatory solutions to mitigate potential harm from the use, re-use, and recombination of these data.

The section on Translational Readiness examines global approaches that seek to improve the integration of novel data streams, namely human mobility data, into response planning by researchers, policymakers, and response agencies. Early efforts to address these challenges are through the socialization of end products, the promotion of data “bilinguals” who can navigate both the science and regulatory realms, the creation of regional hubs, networks, and multi-stakeholder “assemblies,” and direct training and capacity building within response agencies. Incentives to use these data within academic institutions, governments, non profit organizations, and technology companies are nascent and unaligned, despite the potential for analysis to be of use in a disaster or public health emergencies. More evidence on the utility of these data is needed to improve incentives. In order to develop such evidence and enhance the approaches for integrating novel data streams, more data sharing is also needed.

The “Data-Methods-Translational Readiness” framework presented in this paper brings together key issues around the preparation of data for timely use, applying the data meaningfully and purposefully, and nurturing local capacity to receive and act on the analysis. The paper presents a broad view of the state of the art and lists key domains of inquiry to be pursued by technology companies, scientists, lawmakers, and response agencies for the responsible use of novel data streams to maximize public good without causing or exacerbating harm.

We thank Abshishek Bhatia for synthesizing the notes and insights from the seminars that are featured in this document, Sraavya Sambara for her research support, Navin Vembar and Nishant Kishore for their technical inputs, and Joseph Nallen for editing, designing, and distributing this white paper.

We thank Maham Khan and Takahiro Abe for diligently documenting the seminar deliberations in their role as rapporteurs, and Nick Jones for facilitating GFDRR's participation and organizational support. We also thank the Radcliffe Institute for Advanced Study at Harvard University for supporting the seminar series.

We are deeply grateful for the seminar participants' and featured discussants' time and valuable insights, and for making the seminar series possible in the first place. Furthermore, we appreciate their feedback and suggestions as finalized in this white paper.

Sincerely,

**Satchit Balsari, Caroline Buckee, Jennifer Chan,
Andrew Schroeder**

April 2022

Data Readiness

The types of data required to inform critical public health responses in emergencies – including data on medical vulnerabilities, healthcare infrastructure, shelters, environmental conditions, and human mobility – now exist in vast quantities, are up-to-date, and are mostly digital. However, these data reside in silos across various industries and their use by public health responders is limited by technical and regulatory barriers to access and their limited “fitness for purpose.” Additionally, the lack of standardization and interoperability of these novel data streams limit their utility in emergency response contexts.^{8,20}

In this section on Data Readiness, we examine the technical, regulatory, and ethical barriers to accessing human mobility data generated from cell phones. The discussion explores how access is negotiated between the companies who claim ownership of the data, the communities from whom the data are generated, and the researchers and policymakers who seek to use the data. How easily the data are shared depends not only on the technical capacity to send or receive them but also on the privacy and security implications of sharing the data.^{8,20} The lower the resolution of the shared data, the harder it is to breach privacy. This impulse to aggregate and anonymize data is, however, in tension with the need to apply high-resolution data in public health contexts. For example, an aggregate movement vector observed during a wildfire event in California may show movement towards the fire and not away from it. Disaggregated data may show very different temporal patterns, where there is first movement away from the fire, then an inward rush of first responders, followed by communities returning to their homes a few days later. These patterns are revealed when data are available in smaller time bins. In the absence of laws regulating the use of these data, how much data are shared, in what form, and for what purpose is determined by private corporations. The resultant proprietary solutions are inadvertently nonuniform, creating significant interoperability challenges. We conclude by describing legal and societal pathways that seek to change how access is arbitrated.

Provision of Access

This white paper addresses the application of human mobility data generated from cellular devices, either via Call Detail Records (CDR) from telecom companies or GPS traces from mobile applications. Researchers have used CDR data to estimate population movements for about a decade.²¹⁻³⁰ Historically, CDR data have been processed and analyzed internally by telecom corporations for commercial purposes, such as the evaluation of mobile network operators (MNO) markets, advertisements, and monitoring churn.⁷ Early private-academic partnerships seeking to use human mobility data to estimate population movements included bespoke and often drawn-out nondisclosure and data use agreements; and the outputs allowed were limited in scope and general accessibility, but nevertheless laid the foundation for epidemiological applications.⁷

The sharing of CDR data is strictly controlled in many countries by an independent regulatory body. Providing it for research or public health represents a cost to the operator in terms of personnel required to do the processing. Additionally, companies are often reluctant to share data that can be seen as controversial by the public, and therefore access is limited.⁷ The contracts between academic institutions and telecom companies have been resource and time intensive, precluding new relationships or agreements amidst an emergency. During crises, CDR data have been available primarily through pre-existing and trusted relationships.³¹ Even during the early stages of the COVID-19 pandemic when population mobility data were considered critical to understanding and informing non-pharmaceutical interventions impacting travel and mobility, CDR data remained inaccessible to many scientists, policymakers, and practitioners in most countries. There was one notable exception: under pandemic-related emergency provisions, several governments seconded CDR data to track not just populations but also individuals.^{32,33}

In recent years, with the exponential growth of smartphone ownership around the world, individual and population movement has been tracked via GPS signals from social media platforms and other mobile applications that have access to the individual's location. The use of aggregated GPS traces from social media platforms and other mobile applications, in



contrast to CDR data, remains largely unregulated.^{34,35} With the exception of the European Union's General Data Protection Regulation (GDPR), the California Consumer Privacy Act, and other similar regulations, most regions have few laws governing the third-party or public use of aggregated personal or demographic data.³⁶ Many have argued that regulations like GDPR, while protective of individual privacy, hinder access even for purposes deemed to be of a societal benefit. In the absence of regulations, data access is mediated through individual relationships between technology companies and paying clients or non-paying research partners and response agencies.³⁷⁻⁴³ Throughout the pandemic, a growing number of technology partners have begun to release aggregated, anonymized datasets through data use agreements that are available publicly or through low-friction eligibility criteria to a wide pool of interested third parties. Technology companies, like their telecom counterparts, decide with whom they share their data. The Data for Good program at Meta (previously Facebook Data for Good), for instance, has a two-tier approach. Many of its datasets are shared publicly with a range of government, nonprofit, and commercial actors. Other controlled access datasets are shared with on-boarded nonprofit organizations and research groups, of which there are currently 550 across 70 countries. These data use agreements, unlike Creative Commons licenses that govern the use of its publicly available data, often require bespoke negotiations to fulfill legal safeguards at all collaborating institutions.

Consequently, researchers have expressed an interest in developing a common, shareable framework in lieu of the drawn-out data use agreements that would be acceptable by technology companies and legal reviewers at collaborating institutions. In response, a collaborative effort led by the United Nations Sustainable Development Solutions Network (SDSN), New York University's Governance Lab (GovLab), the University of Washington, and the World Economic Forum known as the Contracts for Data Collaboration (C4DC) has gathered and analyzed example data sharing agreements (DSAs) that have been used to share MNO data for health applications to help guide other data actors considering similar arrangements. By collating a repository of vetted and commonly accepted modular contractual clauses, C4DC seeks to demystify the complexity of data sharing agreements. Its services

seek to reduce transaction costs and “strengthen trust, transparency, and accountability of cross-sector data collaboratives.”⁴³ It has also been argued that licensing agreements should instead be replaced by openly available differentially private data. We discuss the limitations associated with this approach in subsequent sections.

The Mechanics of Access

Data sharing practices vary widely across private sector actors. They encompass the sharing of raw Excel spreadsheets with Personally Identifiable Information (PII), to secure role-based access and remote querying. In our landscape, we observe that the entire range of practice persists. Most technology companies allow secure, authorized, role-based exports of select anonymized datasets, often with a human in the loop. API-enabled queries are less common. Some companies will retain data and permit remote querying while others work closely with each partner to release a heavily restricted, fit-for-purpose, single-use dataset. The resolution of the shared data and the associated risks and liabilities determine the ease of access. The following non-exhaustive list of examples illustrates current industry practices.

Cuebiq

Cuebiq is a location intelligence and services provider for a wide range of mobile applications. As such, it collects a significant amount of spatially and temporally granular data on human mobility. While Cuebiq has historically maintained a mobility data publication pipeline for aggregated statistics at the county or census tract level in the United States through Amazon Web Services (AWS), its most recent efforts focus on enabling secure and private sharing of device-level data through the Workbench platform. Workbench, which is accessible only through data agreements with approved researchers and nonprofit organizations, allows an authorized individual to use their own models and scripts through a secure Jupyter Notebooks environment. While raw GPS trace data remains in Cuebiq servers, analytical products at approved scales of resolution can be derived and shared. Cuebiq data using differential privacy was used by the University of

Toronto and the Institute for Scientific Interchange (ISI) foundation to highlight human mobility change and their proximity during the COVID-19 pandemic in Canada.⁴⁵ Of note, Cuebiq was not set up for sharing their data with many research partners in early 2020 but responded to the growing demand for their data during the pandemic.

Footnote: As of May 2022, the location data procurement business at Cubiq has been renamed Spectus.

Data For Good at Meta (previously Facebook Data for Good)

Data for Good at Meta offers several platforms through which it shares data. For public datasets, such as its high-resolution population density maps, Movement Range Maps, and the Social Connectedness Index, the team maintains instances on Humanitarian Data Exchange (HDX) and Amazon Web Services (AWS). In 2021, its publicly available datasets were downloaded over 240,000 times from HDX, and the high-resolution settlement layer was accessed 4.8 million times from AWS. Privacy is protected in these public datasets through aggregation and noise addition, and in some cases, calibrated to meet the standards of differential privacy.

For controlled access datasets, the team offers a centralized portal to disseminate data on granular mobility, mobile network coverage, long-term displacement, and other datasets. These data are shared via Meta's Data for Good Partner Portal for authorized users such as universities, nonprofits, and UN agencies. Interested parties need to apply to become an authorized user by contacting the team and providing background on their organization and associated research or programming. A user agreement needs to be signed by the interested parties. Users can utilize the data available from the Portal to fulfill their research and action purposes but are prohibited from sharing the data directly with a third party; however, insights and derivative products such as resulting research papers and other analysis may be shared. Users can directly download the data in the format of CSV files and GeoTIFF images. The storage of data for a specific event may expire due to data retention policies.

Throughout 2021 and 2022, an increasing percentage of Meta users have discontinued the sharing of their location history data, one of the primary data streams used by Meta's Data for Good program. This is largely due to changes in mobile phone operating systems, which now ask users to re-confirm or re-enable their data sharing preferences on a near weekly basis. As a result, location history data have become more thinly distributed across Meta's user base and the data lacks the granular clarity it once had. Consequently, data products that rely heavily on fairly complete trajectories of movement patterns, such as those used in emergency response or crisis contexts, will be impacted.

Meta has shared its plan to replace pre existing location history datasets with location services versions, as this may help circumvent the challenges associated with updated operating system features and will provide more robust population and spatial coverage. However, updated data will be of lower resolution in terms of tile size – no tiles smaller than 2.4km will be accounted for – and will be updated less frequently – datasets will not be updated more frequently than on a daily basis. While switching to Location Service Data will likely increase the data's representativeness, the decrease in spatial resolution may pose significant challenges in producing data products and maps to support crisis response efforts in small territories and islands.

These proposed changes to Meta's data policies affect around 10 of their 35 program datasets but disproportionately affect those most commonly used in crisis response. The policy changes are planned to go into effect by May 31, 2022. After the changes go into place, some movement range metrics will likely not be recoverable and won't have replacements.

Telenor and other MNOs

As a mobile network operator (MNO), Telenor has historically partnered with academics and researchers to identify a key question of public health interest and provided bespoke aggregations for analysis through a private and secured file sharing platform. These partnerships have been facilitated by Telenor Research based at the company headquarters, which has served as a buffer between tower-level data provided by

business units around the world and researchers. In the past, tower-level data have been provided to Telenor Research, which has worked with academic partners to aggregate the data further to specific administrative units or spatial grids, and on a particular temporal scale. Secure access to these secondary data sets is then provided to researchers.

Other MNOs have made data available in different formats, including providing tower-level information about the number of subscribers moving between towers daily or within particular time windows. In 2014, Orange made specific data sets from West Africa available to teams from the research community as part of a “Data for Development (D4D)” challenge. However, mobile operators typically do not save CDRs indefinitely due to stipulations included in their spectrum licenses, and studies have primarily been retrospective and constrained to a particular window of time, despite these aggregation protocols being used to provide the government with daily reports during the pandemic.

Google Mobility Data

Google’s COVID-19 Community Mobility Reports were developed as a part of Google’s broad efforts to help fight the COVID-19 pandemic.⁴⁶ The mobility reports can be readily downloaded as PDF and structured CSV files. No extra application or authorization steps are required. The data became available for the public in mid-2020 under the protections of differential privacy. The mobility reports are generated at the global, national, and sub-national levels to illustrate how foot traffic has changed at different scales and in different venues or points of interest compared to the pre-pandemic baseline. According to their website, these reports will be available for a limited time, “so long as public health officials find them useful in their work to stop the spread of COVID-19.”

The Humanitarian Data Exchange

The Humanitarian Data Exchange (HDX), managed by the United Nations Office for the Coordination of Humanitarian Affairs’ (OCHA) Centre for Humanitarian Data, has served as “an open platform for sharing data across crises and organizations.”⁵⁰ Launched in July 2014, HDX enables the sharing of data about the

context in which a humanitarian crisis is occurring (e.g., baseline/development data, damage assessments, geospatial data); data about the people affected by the crisis and their needs; and data about the response by organizations and people seeking to help those in need of assistance. HDX currently hosts more than 18,000 datasets which were accessed by over 1.3 million people last year. In 2021, Meta’s Population Density Maps were downloaded about 100,000 times, under a Creative Commons international license.^{9,51,52} Data sharing by private sector organizations remains relatively limited; out of 300 contributing organizations on HDX, only a handful come from the private sector.⁸

Tracking Population Density Changes During the War in Ukraine

When the Ukraine war broke out on February 24, 2022, the CrisisReady team produced map analyses showing changes in population densities in-and-around the Western border of the country using anonymized and aggregated human mobility data from Meta. These maps indicated changes in population density and mobility on a daily basis during a moment when the influx from Ukraine was producing a particularly strong signal of population change.

The analyses did not necessarily indicate forced displacement, and could not be used to measure total numbers of refugees in a particular area, given that there was no ability to subset the data according to Ukrainian origin. However, the rates of change in mobility did allow for quantification and illustration of anomalous mobility patterns which indicated probable differences in refugee flows over space and time.

These maps have been used by international response agencies, including UN OCHA, UNICEF, IFRC, Mercy Corps, and IRC, to highlight directional signals of population movement patterns related to the Ukraine war. The role of these maps is to augment or supplement existing operational updates, to help align local, regional and national response efforts, to support decision-making at the operational and tactical levels, and to aid the allocation and deployment of resources for those displaced by the war.

Resolution vs. Utility

As described in the examples above, human mobility data are typically aggregated and anonymized by technology companies before they are shared with researchers. The granularity of these released data — particularly the temporal and spatial scales at which they are released — are calibrated to protect the privacy of individuals and groups, and noise is often added in order to achieve differential privacy. With no industry standard in place, companies choose different methods of aggregating data and introducing noise. This includes excluding data where there are fewer than ten individuals, temporal aggregation, and the inclusion of noise into metrics of interest, making the data differentially private.⁵³⁻⁵⁶ There is growing recognition that this approach may be suboptimal for several reasons:

1. On the one hand, protecting personal identity may not be sufficient, as demographic or group identities may be revealed by looking at aggregated patterns, placing entire groups or neighborhoods at risk.⁵⁷ Data on mobility to-and-from sensitive points of interest like abortion clinics, gay bars, or mosques may place both individuals and groups at risk.
2. On the other hand, low temporal and spatial resolution may lower the public health utility of the data. Population dynamics in densely populated areas, for instance, may be misrepresented by how aggregations are applied using larger tile schemas. A large city, such as Mumbai, would require relatively more granular and higher resolution spatial scales to effectively represent mobility patterns, as compared to a low density rural area, which would require larger tiles and lower resolution to increase sample size and control for the inadvertent release of privacy-protected information. It is worth noting that during the COVID-19 pandemic, while researchers sought out higher resolution data, public health officials did not necessarily need it as much as they needed actionable insights.
3. In general, information on the representativeness of the data vis-a-vis the general population is

difficult to determine. The amount of noise added to the aggregate counts (e.g., to satisfy differential privacy) and the associated uncertainty are often not precisely specified or uniform across companies. This adds uncertainty that is difficult to quantify in the downstream analysis using the data. Safeguard methods like differential privacy, although more rigorous, are often difficult to explain and communicate to end users as to exactly what kind of protection they receive.

The Salience of Context

The safeguards that disappoint researchers preclude downstream abuse by malfeasant state or non-state actors. The question then is how to determine the granularity of access. Who decides? For a long time, it has been recognized that the answer is context-dependent.⁴⁴ Raymond et al, argue that there are some categories of data, such as sexual orientation or religion, that must not be touched in most contexts: unhand them (*noli mi tangere*). There are other sets of data, and consequently levels of resolution, that may be shared with certain partners under specific circumstances, as long as they do no harm.⁴⁸ What may be appropriate for sharing with a public health agency may not be for law enforcement.^{58,59} Mobile phone data are often requested by law enforcement agencies, diminishing trust and participation in data collection or data sharing endeavors for public health purposes. For example, after Singaporean citizens found out that the government-endorsed contact tracing app TraceTogether was feeding data into law enforcement systems, there was public outrage and a consequent decrease in trust in the federal government.⁶¹ During an exploratory discussion about the use of digital contact tracing apps in Massachusetts, several civil society representatives and human rights advocates pointed out that racial minorities and immigrants, some of the more vulnerable groups during the pandemic, were also those least likely to participate in the system due to low levels of trust.⁶² The COVID Mobility Data Network's (CMDN) researchers working with public health agencies in a major East Coast city were specifically requested not to share population mobility data with law enforcement agencies. South Korea and Taiwan, whose public health agencies and governments

engendered high levels of trust from their populations, witnessed greater public participation. As the Inter-Agency Standing Committee (IASC) Operational Guidance on Data Responsibility in Humanitarian Action lays out, data use must principally abide by the “do no harm” principle (*primum non nocere*), before it can try to achieve a net benefit.⁶⁴ While the breaching of privacy is considered inherently problematic, the harm or exploitation that could come from the privacy violation is perhaps even more relevant to determine whether the data should be shared.

Currently, these decisions are made on a case-by-case basis and vary across the world based on social norms, political pressure, institutional expediency, and legal imperatives.⁶⁶ Throughout the pandemic, state agencies and researchers requested access to mobile data while they had variable capacities to responsibly receive or safely store, process, or destroy the data. Whether or not data should be shared at a particular resolution, Raymond et al propose, should also be dependent on the receiving ecosystem’s Capacity, Capability, Competency, and Culture – issues we discuss in the final section on Translational Readiness. These 4Cs complete the prerequisites to data access:

1. Can these data be touched?
2. Can these data cause harm?
3. Does the receiving entity have the capacity, capability, competency, and culture to use these data responsibly and safely?

Consent and Its Limitations

In the absence of regulatory frameworks that allow easier access to human mobility data, technology companies are currently the sole arbiters of access. As discussed above, whether datasets will be created and shared, with whom, and for how long, is almost entirely determined by the data companies, albeit with inputs from collaborators in government, research, and nonprofit communities. The growing global alarm around the use of personal data collected actively and passively on mobile devices led to a spate of corporate policy changes in 2020 that promoted (and

necessitated) individual consent for the secondary use of data.⁶⁵ While this did not make the data more readily available to third-party users, it provided corporations the legal coverage to use the data and reportedly gave individuals the option to not share their data. The App Tracking Transparency (ATT) framework introduced by Apple now requires user authorization to access app-related data for tracking the user or the device. Google has now blocked un-consented third-party tracking in Chrome. These approaches are not dissimilar to the traditional use of “bedside” consent in medical practice and clinical trials.

However, privacy advocates have long argued that such consent is insufficient due to a variety of reasons, including the often coercive nature of its collection, the questionable “informed” nature of consent given the impossibility of predicting or imagining future use, and the unintended risks or harms caused by the inadvertent combining and recombining of big datasets.⁶⁶⁻⁷⁰ India’s proposed Data Empowerment and Protection Architecture (DEPA) framework is a regulatory solution that seeks to address these limitations of consent on downstream use by permitting consented access to a recursive use of high-resolution data across multiple partners. Driven by open standards, the DEPA framework would permit the secure, audit-able re-use of data through notification, where the data principal retains the right to revoke access or to “opt-out.” The act of consenting is temporally removed from the actual porting of data, providing data principals opportunities to make informed decisions.

While such consent-driven time and purpose limitations – concepts derived from the FAIR (Findable, Accessible, Interoperable, Reusable) Principles and GDPR – still drive the downstream use of these data, many consider them outdated and insufficient.⁶⁰ Often these data contain not Personally Identifiable Information (PII), but Demographically Identifiable Information (DII), which we have stated can also be problematic, or simply, Action Based Information (ABIs), as described in previous sections. The concepts of agency and autonomy (as well as consent) are rooted in the privacy and dignity of the individual and based on Nuremberg Principles that continue to influence most health data protection frameworks, including the Health Insurance Portability and Accountability Act in the United States.⁷¹ These approaches, which also

undergird the GDPR, are not adequate. The subject of human mobility data is not a distinct individual but is often probabilistic. Human mobility data are a mosaic of interstitial and missing data that create a partial probabilistic cohort of individuals or demographic groups that do not lend themselves easily to the predominant ethical conceptions in medicine and public health data management. And yet, in academia, Institutional Review Board (ethics approval or sometimes, exemption) is needed for the use of human mobility data, requiring purpose specification.

The consent-heavy restrictions on AdTech data also preclude society from benefiting from advances in machine learning and data science.⁷³ What then are the ethics of not using such data for public good? An innovative legal tool from India, the proposed “Non-personal Data Protection Bill,” takes this stance. It seeks to mandate access to de-identified datasets deemed to be of public value. The Bill proposes that all data businesses be mandated to publish their metadata for review by community-level data trusts that determine whether the dataset is of high public value. Those deemed to be of high value must be shared with designated data trustees (governmental or nonprofit organizations), who are responsible for the maintaining and sharing of these data with public and private organizations for public good.⁷⁴ Were technology companies compelled to share such data for public good, what mechanisms would protect individuals or groups from state surveillance?⁷⁵

To overcome the presumed incompatibility of robust anonymization and data utility and move beyond consent, data scientists have proposed and begun to apply differential privacy, which guarantees that individual-level data won't be revealed, provided that noise is appropriately introduced into calculation of published aggregates. The application of differential privacy to the 2020 United States Census is a substantive advancement in how census data are shared. The National Conference on State Legislatures describes the tension between privacy and utility we previously observed with human mobility data: “The dual requirement for an accurate count and the protection of respondents and their data creates a natural tension: The more accurate (and therefore usable) the reported data is, the easier it may be to identify individual responses. And yet, as the raw data is

altered before being reported (to protect confidentiality), the less usable the publicly released data is.”⁷⁶ The amount of statistical noise that will be introduced will depend on the risk to privacy, where publicly shared data on rural populations, smaller households, and racial minorities are likely to have more relative noise introduced due to the low original aggregated counts. OpenDP, an open-source community that develops differential privacy tools, now includes teams examining the impact of introducing statistical noise in human mobility datasets on their epidemiological utility.⁷⁷ While these approaches are promising and have been used by Facebook, Google, and Cubiq for sharing mobility data, there are not yet sufficient public health use cases to help determine their efficacy.

Negotiating Access

The technical mitigation measures proposed above, including the DEPA framework or differential privacy, are in early conceptual and implementation stages. Until there is evidence of work, a less automated approach is warranted, one which recognizes the particular context in which data are applied and the associated risks or harms. A universal, automated model is unlikely to succeed given the vast disparities in infrastructure, social norms, attitudes, technological capabilities, legal frameworks, and enforceability around the world. To advance the theoretical notion of a multi-pronged approach to resolve this tension between privacy and utility and accommodate for societal context, Verhulst et al. have proposed the creation of Data Assemblies. Data Assemblies solicit “diverse, actionable public input on data re-use for crisis response in the United States.”⁷⁸

The first Assembly in New York worked with civil rights organizations, key data holders, policymakers, vulnerable communities, and the public to co-develop a responsible data re-use framework to inform the re-use of personal data. The Gov Lab argues that signals about social attitudes and values toward data use come from newspaper opinion pieces and surveys, which may (or may not) be biased, and which offer a specific snapshot of public opinion but tend to lack the nuance unearthed by more deliberative methods.



Multi-stakeholder discussions help the group understand how various constituencies perceive risks and value. Empowering end users to understand the potential of these data and imagine ways to embed them into their decision-making algorithms is likely to result in more purposeful applications. Such collaborative approaches would result in “pulling only needed data” to achieve specific tasks.

There is precedent for the sharing of private data during emergencies at larger scales, as well. The International Charter for Space and Major Disasters, for example, facilitates the sharing of satellite imagery by commercial providers, on request and under a specific set of triggered conditions.⁷⁹ The recipients are familiar with the data and receive it at a pre-set, pre-negotiated level of granularity. Such approaches are not without implementation limitations. The Charter could not be activated during crises in Nepal and Haiti, as those nations did not have the political leverage to do so. As a workaround, data was sought from the National Geospatial Agency, through the U.S. State Department.

It is unlikely that the inclusive, participatory approach promoted by Data Assemblies will quickly reach the vast scale envisioned by automated consent manager systems, like DEPA. However, their deliberative stance makes consent more meaningful and the use and the re-use of data more purposeful. As the Assemblies do not have the power to mandate data from the technology companies, they would have to make a case for the use of these data to a variety of stakeholders. To date, most technology companies make these data available as part of their corporate social responsibility programs, which make substantial investments in time, personnel, and infrastructure to develop, maintain, and disseminate meaningful, secure and anonymized datasets. Therefore, there is palpable pressure on the “data for good” programs embedded in large data companies to provide and demonstrate value. While there is precedent for sharing other private data (like clinical or lab data) during disasters, the sharing of human mobility data is relatively nascent. Without the results of a substantial amount of pending methodological work by scientists and policymakers, it is not easy to demonstrate value based upon lessons from implementation in both the scientific and gray literature. To justify the utility of these data, scientists need more data. To provide more data, the companies

want to see proof of its impact. In the next section, we discuss the methodological work that needs to be prioritized to advance the application of human mobility data in disasters.

Action Points

- 1. Introduce legislation to govern the use (and re-use) of such novel data streams, including human mobility data.*
- 2. Launch a professional body comprising technology companies, researchers, and response agencies to publish interoperability standards on human mobility data and other novel data streams.*
- 3. Generate context-specific consensus on aggregation and anonymization of shared data.*
- 4. Promote the development of standard contractual language for the use of human mobility data by academics and policy makers.*
- 5. Promote a cadre of data stewards within technology companies to guide the responsible and meaningful use of the data for public good.*
- 6. Develop, test, and disseminate use-cases on the application of differential privacy on human mobility data sets for epidemiological or other public health purposes.*
- 7. Include communities from whom the data are generated in defining the scope of use of data.*

Methods Readiness

The expedient application of human mobility data during a public health emergency, even when accessible, is often impossible. Unless researchers and policymakers are already acquainted with the provenance of the data and the opportunities and challenges of their application, multiple barriers appear to prevent the utilization of these potentially useful data. These barriers can include relative novelty of human mobility datasets, the proprietary (and variable) nature of data processing, and variation in analytic techniques among research groups delays.⁸⁰

Several methodological issues are associated with their use. The data's representativeness and the potential harm that may stem from their use are identified by this report as being the two key issues. As discussed above, calibrating the resolution of the data shared along these two axes are determined by the legal, societal and political context in which the data are used. It is also determined by the presence or absence of a theoretical foundation for evaluating the potential contextual appropriateness and efficacy of available data.

Understanding Bias and Estimating Representativeness

Human mobility data is an output of the behavior of device owners. An important consideration for any data set is the market share or coverage of the operator. CDR data from telecom companies, which represent the subset of the population that subscribes to a particular mobile service, may skew toward a certain geographic, economic, or social demographic.^{81,82} These contextually specific demographic disparities in what cohorts can and cannot access certain carriers, let alone mobile devices in general, must always be considered when evaluating how representative a CDR set may potentially be.

Further, routine maintenance and shifts in tower locations can cause anomalies in CDR data that require constant communication with the data scientists and

the operator. In many countries in the Global South, it is common practice to switch subscription plans, SIM cards, or phone numbers, or to share a mobile device among several household members.⁸³⁻⁸⁶ This means that analytic frameworks need robust contextual knowledge both from the operator itself and about the region or country the data comes from. Data from AdTech companies will represent users of particular kinds of apps, which may, in turn, be biased by age, gender, socioeconomic status, religion, political beliefs, and so on. While the bias in data determines the uncertainty associated with the analysis, in the context of the human mobility data, representativeness is often closely related to (un)fairness and (in)equity. If the data tell the stories of people who own these devices, whose stories remain untold? In the case of the pandemic, the economically disadvantaged, children, and the elderly may be the very populations that did not have access to Bluetooth Low Energy (BLE) -enabled smartphone devices, a critical component of digital contact tracing efforts. In communities that made digital contact tracing the cornerstone of their early public health response, those most at risk often had the least access to the services.⁸⁷

In the case of human mobility data, say for forecasting, observing and responding to population evacuations following natural disasters, reliance on data from smartphones risks missing large swathes of the most vulnerable, such as refugees.⁸⁸⁻⁹⁰ Within refugee populations that have some degree of connectivity, there are other barriers to connectivity access existing within the displaced population, most notably pre-existing gender disparities. In a Harvard Humanitarian Initiative study from 2017, 94% of Syrian men in the Ritsona refugee camp in Greece had access to a mobile device versus 67% of Syrian women.

Another example of this phenomena is the Bidi Bidi refugee settlement of Uganda. There, cell phone ownership among men is two times that of females, introducing a significant gender bias in the analysis.⁹⁰ Most working women in the Rohingya camps in Bangladesh, on the other hand, owned a mobile phone (though not a smartphone), resulting in very different analyses from GPS based datasets as compared to CDR based ones. Understanding local context becomes critical. As technological access follows pre-existing social determinants, data collected from vulnerable



populations must be explicitly analyzed with cognizance of these commonly recurring disparities.

Given the opacity around the representativeness of human mobility data sets compared to the population baselines, researchers often request more information from technology companies typically reticent to share this information in order to avoid. Though there remains a red line around this knowledge, there is growing recognition among technology companies that such information may be a prerequisite for meaningful use of their datasets. In an effort to bring data representativeness into the conversation, Meta has attempted to make users of its data aware of the shortcomings in sample representativeness and encouraged them to use multiple data sources in conjunction.⁹¹ Additionally, Data for Good at Meta has recently undertaken a complex re-weighting of several of its datasets to ensure better population representativeness and will be working to incorporate this methodology into future tools.⁹²

Data streams during crises are often unreliable due to breakdowns in infrastructure and normal processes.⁹³ Power outages, for example, may cause mobility data to be missing altogether for the most affected geographies, and when movement patterns are also disrupted, this may make it difficult to use imputation to make up for data gaps. Other kinds of disruption may also impact data being used in conjunction with mobility data. Inability to keep up with medical charting when healthcare systems are overwhelmed typically results in incomplete and missing data at the peak of crises, for example. Analytic precision is further diluted when multiple data sets with varying uncertainty are combined during disasters. For example, during the pandemic, human mobility data were often used in conjunction with COVID-19 “testing” data representing epidemiological case trajectories in a particular location. Depending on the time and place, however, clinical data encompassed various kinds of antigen and PCR tests with a range of sensitivities, antibody testing, and a dynamic and evolving criterion for testing. In this case, the epidemiological data suffered from different biases and coverage compared to the mobility data, making it difficult to rigorously estimate and communicate uncertainty in analytic products requiring both (e.g., estimates of how the disease may spread geographically). Combining big data streams can both

compound and compensate for missing data. What degree of bias (and therefore, uncertainty) is acceptable is therefore hardly generalizable and closely linked to the application of the data. In some cases, the bias may significantly distort public health decisions, in others, it may have little to no operational impact.

Quantifying Harm

Attempts to quantify harm are focused on the probability of unmasking individual or group identity at different levels of data aggregation. Differential privacy, for example, provides a quantitative framework that allows researchers to probabilistically measure how much more likely an individual is to be re-identified given the availability of a data set. This is a complex field that has many subtleties, which may impact the design of data that researchers release. As discussed above, the loss of privacy is not always the harm, it can be exploitation. There are political and societal contexts when the loss of privacy may not result in harm, as in when governments enjoy high levels of trustworthiness, or in the presence of strong, enforceable legal or social deterrents. As we saw during the pandemic, there was a greater willingness to share data (as with contact tracing) for the perceived benefit. There can, however, be a significant gap between perceived and observed benefit: human rights advocates are justifiably alarmed by the privacy intrusions resulting from digital contact tracing efforts, almost all of which failed to result in demonstrable public health gains.⁹³⁻⁹⁵ Some argue that efforts to apply these data post-hoc to public health problems may merely be seeking to whitewash the otherwise overly intrusive collection process. Incremental gains for science must be balanced with societal expectations and the potential for harm. Societal tolerance for loss of privacy (and potential harms) is intimately related to the danger at hand and can change over time. A benevolent government can be replaced by an authoritarian one, even democratically. Attempts to quantify harm from such datasets cannot, therefore, confine themselves to the mere masking or unmasking of identity.

New York University's Governance Lab (GovLab) proposes the development of a cadre of Data Stewards in technology companies that can help guide company policy on the sharing of data by understanding the

context in which the data will be used. Even as companies move toward providing platform solutions, there may be opportunities for more bespoke arrangements where greater data granularity is calibrated to specific contextual risks. We discuss this further in the context of Data Assemblies and India's non personal data protection bill.^{96,97}

Standardization

While technology companies and researchers recognize the need for standards in accessing, sharing, and applying human mobility data in different contexts, there has not yet been a systematic or coordinated effort to do so. Standard information on representativeness and bias, based on population characteristics, as well as transparency about noise introduced by processing algorithms, are considered critical elements that must be provided to improve the generalizability and translation of data for public health. There are proposals for technology companies to publish metadata along with guidance on associated bias and uncertainty. Scientists and technology companies will need to co-develop validated bias correction and post-processing or analysis strategies to improve the precision and utility of these large datasets. There will, however, be limits to what technology companies are willing to modify or invest resources in. Standardization in temporal and spatial scales across datasets, while desirable, is harder to achieve because of the lack of incentives, unless scientists can make a stronger case for not changing the status quo. The Open Geospatial Consortium's Moving Feature Standard deals with not only human mobility, but all moving features, promoting interoperability and integration of mobility data from different sources.⁸⁵

There is precedent that can be adopted from related industries as well. The Health Level Seven International (HL7) health-care standards organization created the Fast Healthcare Interoperability Resources (FHIR), a standard describing data formats and elements, as well as an application programming interface (API) for exchanging electronic health records.⁹⁹

Combining Mobility Data with Epidemiological Data

The use of mobility data as an input in spatial epidemiological models has been developing over the last decade.^{11–20} Before mobility data of this kind was available, the lack of broadly available information about travel patterns, particularly on spatial and temporal scales relevant for disease transmission, was a major hurdle for producing accurate models of epidemics, for example.²² There have been a variety of studies that combine mobility data with the prevalence of a disease, including malaria, dengue, chikungunya, or measles, to understand how human movement patterns contribute to the spread of outbreaks between populations.^{15–19,100,101} In general, these models are most robust for epidemic pathogens and for diseases impacting urban populations because the epidemiological model is more tractable in the former case (endemic pathogens are challenging to model because there is no clear “wave front” of cases) and the mobile data is usually more robust in urban areas with respect to the latter.¹⁰² New approaches include combining spatially explicit pathogen genomic information – with phylogenetic estimates of connectivity between locations – with mobility data.^{100,103} In 2020, this work was rapidly expanded and applied to COVID-19 all over the world.¹⁰⁴ Modelers used mobility data to predict where SARS-COV-2 would travel to next, estimate the level of transmission (via the reproduction number) in different areas, and quantify the effectiveness of interventions, such as lockdowns and travel restrictions.^{104–109} Because the lockdowns were so universally applied and so dramatic, there was a strong relationship between mobility patterns measured using mobile phone data and the trajectory of the epidemic.¹⁰⁹ This relationship weakened over time as heterogeneous interventions were implemented, and the contact patterns driving disease became decoupled from the travel patterns on larger scales.¹ The COVID-19 Mobility Data Network worked with public health officials to monitor the impact of interventions around the world and to provide epidemiological contexts for the data. Issues that have arisen repeatedly in this literature are how well the spatial resolution of the epidemiological data matches the mobility data and how to manage the uncertainty associated with biases in each data stream. Due to the



urgency of the pandemic and the relative novelty of these data streams within a wide range of academic disciplines, the mobility data were not always used appropriately, reflecting the need for systematic analyses of how and when these new data streams should be used during an epidemic.

Action Points

- 1. Develop frameworks for communicating bias and uncertainty while publishing and communicating analyses using mobility data from mobile phones.*
- 2. Advance methods to address bias correction in human mobility data sets.*
- 3. Support the development of standards to promote interoperability among the data sets, especially across temporal and spatial scales.*
- 4. Develop approaches to allow for the diverse anonymization techniques used across technology companies.*
- 5. Develop a framework for quantifying the potential for harm that acknowledges societal and political context.*
- 6. Advocate for donors and national academies to greatly invest resources for translational data science, in preparation for and during emergencies, including allocating emergency funds.*

Translational Readiness

The pandemic continues to reveal the significant challenges in translating science to practice. Response agencies struggle to effectively introduce novel data sources, like human mobility data, into ongoing preparedness, response, and recovery workflows. Borrowing from clinical medicine's established "bench-to-bedside" approach for translating new techniques and technologies into established medical practice, we can better understand and resolve these challenges through a common rubric of "translational" readiness. Among the keys to effective translational practice within medicine are the need to move scientific findings safely, but deliberately, into actual clinical settings to determine applicability in practice rather than in the more controlled space of the lab, and the related need to generate standard workflows which then incorporate successful novel practices into accepted routines. Similarly, the integration of human mobility data into daily applications can be advanced through a process of foregrounded interventions that include careful listening and attention to agency needs, user-centric design, periodic simulations, and training exercises.

In 2020, policymakers and response planners interacted with population mobility data in several ways: They had access to publicly available mobility dashboards and reports periodically updated by technology companies like Google, Data for Good, and Safe Graph; Context-specific and customized reports were often sent to county, provincial, or city offices at regular intervals by technology or academic partners. An example of network collaboration between researchers, policymakers, and response planners is the COVID-19 Mobility Data Network, which customizes and adapts reports and engages in regular communications between these groups to apply and understand these mobility results. When collaborating networks established trust with one another, they were more successful in analyzing, interpreting, sharing, and acting on the results. These policymakers or response planners were able to work with their research collaborators to articulate their needs and priorities and align researchers on the purpose, context, and

utility of mobility data in their work. For example, the Integrated Data team at the New York City Department of Health and Mental Hygiene stated that they most benefited from human mobility data analysis because of their working relationship with researchers with whom they co-developed custom analytic tools and who helped the team address or resolve methodological challenges such as representativeness or missing data.

In many instances, the response agencies had no opportunity to either shape the research question or the nature of the output. On the other hand, lack of data literacy and familiarity with these kinds of novel data streams also precluded policymakers and response agencies from imagining or demanding their use. Complex analyses, even when scientifically robust, are often misunderstood by planners and responders who are unfamiliar with the data and methods.

The translational challenge is, therefore, multi fold. The analysis and products developed are not contextually intelligent and may not speak to the more urgent needs on the ground. Even if they do, they are often too technical for easy adoption by the untrained end user. Finally, the novelty of these data keeps demand low in most parts of the world, negatively influencing incentives to improve data access or advance methods. The COVID-19 pandemic has been a striking example of the inappropriate use and misinterpretation of prediction models, often with dire consequences to entire populations. Improving translational readiness requires purposeful and strategic expansion of supply and demand, while leveraging the limited technical and human resources available.

Supply-Side Expansion

Technology companies have continued to release data and products for "public good" through their Corporate Social Responsibility (CSR) programs. In 2020, Google and Meta published estimates on population mobility in relation to pre-pandemic baselines to help policymakers assess the impact of lockdown directives. In late 2021, launched tools that estimate the service areas and travel times for health facilities by incorporating the mobility of Facebook users.¹¹⁰ Given the limitations associated with some of these platform

approaches described in previous sections, scientists caution against the direct incorporation of such services into response planning without the right kind of associated domain expertise. Therein lies the challenge. Technology companies would argue that they are already diverting resources to produce these goods, and it would be someone else's responsibility to help translate these products into actionable interventions.

Academic scientists have de facto played this translational role during the pandemic but have neither the professional incentives nor financial resources to sustain this role. For example, researchers from the CMDN partnered with public health agencies around the world to regularly share situation reports and help response agencies interpret new analyses coming down the pipeline to make better decisions about policies affecting lockdowns, business closures, and travel bans.¹¹¹ Any attempts to link mobility data with sero-surveillance data required an even more intensive investment of resources to either encourage or dissuade responders from drawing simplistic causal inferences, depending on the context.¹¹² While all these interactions enriched the data producer-researcher-responder ecosystem, they are largely unsustainable due to the misalignment of incentives for each stakeholder. Scientists are not professionally rewarded for translational work. In fact, they are often dissuaded from doing so early in their tenure process.

Any attempts to augment the "supply side" will therefore require allocating financial resources aligned with a professional reward for the research intermediaries. Awareness and interest among philanthropic donors and national academies can be catalyzed through the dissemination of best practices and peer-reviewed material that underscores the utility of these data or the need to address methodology and implementation gaps. Sustained demand by researchers will help drive upstream production, allowing Data For Good programs to demonstrate need within their own corporations. Unlike capital expenditure CSR investments, it is hard to fit the impact of these data in the simplistic, overused, and almost always inflated "Saved X lives" rubric that investors demand.

Demand-Side Expansion

To meaningfully incorporate human mobility data and related data streams into public health decision making, government agencies and organizations need to greatly improve in-house capacity.¹¹³ If academics can demonstrate the effective and responsible application of these data, ideally, governments and the data providers themselves (who are presumably incentivized to have the data be used and useful) should "productionize" those methods. To date, successful use or integration of mobility data into planning or response have largely depended on individuals within government agencies and not on systematic or programmatic mechanisms to do so. For example, in the British Columbia Center for Disease Control, researchers held leadership positions in the public health agencies, which allowed them to successfully leverage mobility data in the government COVID-19 response. While this is natural and expected in the early stages of any novel application or innovation, policymakers are unlikely to engage with the data products produced upstream unless they can recognize their value. High turnover in response units in the humanitarian and public sector frequently disrupts any established momentum and institutional memory of the use of these data (and the lessons thereof).

It is important to educate policymakers about the potential harm that the use of these data can produce, as well as the uncertainty associated with the data. As discussed earlier, simplistic and literal interpretations of human mobility data, especially in relation to lockdowns, were used by law enforcement agencies in many parts of the world to discipline – sometimes with brutal force – individuals or groups who were deemed to be noncompliant, regardless of the context. Whether individuals could afford to stay home, whether they were adequately masked or not, or why they were unable to shelter in place was neither captured in the data nor considered in the knee-jerk response during the early months of the pandemic. A study focusing on Los Angeles County neighborhoods revealed that low-income communities have greater shelter-in-place burdens than more affluent ones.¹¹⁴ These burdens, including the economic stress of job loss, lack of access to personal protective equipment, and sometimes not

qualifying for government relief, had pushed many low-income families to take higher health risks.

Policy makers must also be educated on the uncertainty associated with these data (due to associated challenges with representativeness, bias, missing-ness, etc.). There will be varying levels of comfort with such uncertainty, both conceptually and operationally. Practitioners from response agencies expressed comfort with some amount of uncertainty over no information at all.

Improving human mobility data literacy requires both socializing policy makers to the utility of these data and including policy makers upstream in the data processing and analytic phases. In order to help government agencies calibrate their demand-ask (not too much or too little), needs-finding must consider the question, “What do you want to do?” over the more commonly asked “What do you want to know?” This is an important distinction to make to check possible governmental surveillance. If data are given freely to governments in pursuit of a public benefit, they can be misused by later administrations.¹¹⁵ Corporate transparency about data sharing agreements with governments allows the public to hold governments accountable.

To improve demand and capacity, the Gov Lab has proposed developing a cadre of “data bilinguals,” trained data professionals who understand both the data science as well as the policy or response implications and serve as informed intermediaries driving translational readiness. Their task is as much to communicate the utility of these data to public agencies as it is to recognize the harm and inequities that can arise from the use of these data. They will also be well-positioned to signal information and knowledge gaps from the field to researchers and technology companies to optimize the availability of the right kind of data at the right time. Field practitioners have proposed the integration of data-driven decision making into disaster preparedness drills and simulations. Bilinguals would be well poised to understand information gaps during these simulations and translate them into data and analytic needs for upstream partners. Both supply and demand are likely to respond to public calls for the use of these data. Large media houses, including The New York Times and

The Washington Post, host in-house data scientists and data visualization experts that produced some of the most effective data communication tools throughout the pandemic. Op-eds, media interviews, podcasts, and other dissemination tools directly speaking to the utility and limitations of these data, as was done by supporters of digital contact tracing efforts, will help advance public awareness and drive adoption under suitable circumstances. Governments that do not engender trust are likely to face opposition to the use of such data. The Data Assemblies proposed under the Data Readiness Section, may allow for the thoughtful, context-specific negotiation to occur.

There are several domestic and international examples where institutional efforts have supported the integration of data into response pipelines. The UN's Regional Collaboration Centers (RCCs) support “national climate action through capacity building, technical assistance, and strategic networking – sourcing know-how and resources to drive clean development.” Recognizing the authority of member states and organizations at the table, the RCCs have had success in mediating legitimacy and autonomy for data units embedded in state agencies by lowering bureaucratic barriers. In the early months of the pandemic, the California state government convened an expert group of modelers to serve as a single source of truth for outbreak prediction, help navigate divergent models, and address the high noise to signal ratio.

Based on growing evidence that acting before the onset of a predictable shock is significantly faster, more dignified, and more (cost-)effective than traditional humanitarian response, the Office for Coordination of Humanitarian Affairs (OCHA) has facilitated the setup of multiple anticipatory action frameworks.^{116,117} Each pilot framework comprises three core elements: a) A robust forecast-based trigger embedded in a clear decision-making process (the model); b) Pre-agreed action plans that can mitigate the impact of the emergency and the need for humanitarian relief (the delivery); c) Prearranged finance (the money). Similarly, generating evidence that the purposeful use of human mobility data can meaningfully address public health

challenges requires funding to demonstrate utility, strengthen methodology (as previously discussed) in “peacetime”, and support predesignated rapid response research-policy collaborations during emergencies.⁸

Action Points

1. Document and disseminate best practices targeting:

a. *technology companies, to incentivize them to keep sharing these data responsibly in the absence of a regulatory mandate to do so;*

b. *donors and national academies, to incentivize them to greatly invest resources for translational data science in preparation for and during emergencies. Stronger investments in advancing both the methodologies as well as the local capacity to receive and act on the sights, are required to generate the evidence that the data are useful. During emergencies, provide funds to accelerate translational work.*

c. *policy makers, to socialize them to the potential use of these data products and to the expertise required to generate, process, and analyze them, and to help response agencies guide purposeful, actionable analytic products;*

d. *general public, to improve transparency about the potential use and limitations of these data to promote informed adoption.*

2. Fund and develop a cadre of “data bilinguals” embedded in public health and response agencies.

3. Fund and sustain a distributed network of researchers to support local regional collaborations of trained scientists and response agencies.

Conclusion

Human mobility data hold great promise in informing population dynamics. How people move, when, and where, in anticipation of disasters, during crises, or in their aftermath, has a direct bearing on their wellbeing and safety. Information on population movement can provide critical information to response agencies charged with pandemic preparedness, rescue, recovery, or rehabilitation. Along with longer temporal scales, these data can provide societies information on transient, seasonal, and permanent migration patterns. In malevolent hands, these data can be used to target and harm individuals and entire communities. There is very little legal or technical guidance on the safe, responsible, meaningful, and purposeful use of human mobility data. There are fewer incentives for developing strategies to address these gaps. Current applications of human mobility data are largely driven by data availability, research interests, and to a lesser degree, operational response needs.⁹⁷

The Radcliffe Seminars sought to identify key opportunities to advance the use of these data for public good. We have organized the issues identified around our Data-Methods-Translational Readiness framework to identify technical, legal, and educational priorities for the coming years.

Among the most important lessons learned during the pandemic are the following:

1. Human mobility data can be very useful in informing public health response, but these data have limitations. Among them these are issues of representativeness, bias, and data gaps (missing-ness). Device ownership and individual, familial, and societal patterns impact the data generated from these devices. This contextual information is not typically readily available.
2. The controlled access to these data can be mitigated by pre-negotiated contracts, collective agreements, and better regulation.
3. There is an urgent need to financially support the methodological work that remains on the use of cell-phone-derived location data in public health applications.
4. There is an equal need to accelerate global capacity to use and apply such novel data streams. This need can be somewhat through distributed networks of trained professionals and regional cooperation among local scientists and policy makers.
5. Concerted efforts need to be made to embed data-driven decision making into preparedness exercises to “routinize” the use of these data in crisis response.



References

1. Kishore N, Kiang MV, Engø-Monsen K, et al. Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. *The Lancet Digital Health* 2020; 2: e622–8.
2. Cinnamon J, Jones SK, Adger WN. Evidence and future potential of mobile phone data for disease disaster management. *Geoforum* 2016; 75: 253–64.
3. Wilson R, zu Erbach-Schoenberg E, Albert M, et al. Rapid and Near Real-Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters: The 2015 Nepal Earthquake. *PLoS Curr* 2016; 8: ecurrents.dis.d073fbeece328e4c39087bc086d694b5c.
4. Acosta RJ, Kishore N, Irizarry RA, Buckee CO. Quantifying the dynamics of migration after Hurricane Maria in Puerto Rico. *PNAS* 2020; 117: 32772–8.
5. Unlocking the power of mobile data for humanitarian purposes. <https://www.flowminder.org/resources/publications/flowkit-unlocking-the-power-of-mobile-data-for-humanitarian-and-development-purposes> (accessed Nov 21, 2021).
6. Talia Gerstle, Maya Hardimon, Galen Hunt, et al. Assessing the Use of Call Detail Records (CDR) for Monitoring Mobility and Displacement. 2021.
7. Buckee CO, Engø-Monsen K. Mobile phone data for public health: towards data-sharing solutions that protect individual privacy and national security. *arXiv:160600864 [cs]* 2016; published online June 2. <http://arxiv.org/abs/1606.00864> (accessed Oct 23, 2021).
8. Private Sector Data for Humanitarian Response: Closing the Gaps. *Bloomberg New Economy*. 2020; published online Sept 25. <https://www.bloombergneconomy.com/news/private-sector-data-for-humanitarian-response/> (accessed Oct 23, 2021).
9. Maas P. Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage AK USA: ACM, 2019: 3173–3173.
10. “Facebook Data For Good Movement Range Maps.” Accessed January 31, 2022. <https://dataforgood.facebook.com/dfg/tools/movement-range-maps>.
11. COVID-19 Community Mobility Report. “COVID-19 Community Mobility Report.” Accessed January 31, 2022. <https://www.google.com/covid19/mobility?hl=en>.
12. “SafeGraph | Academics.” Accessed January 31, 2022. <https://www.safegraph.com/academics>.
13. “Covid-19 Social Distancing Scoreboard — Unacast.” Accessed January 31, 2022. <https://www.unacast.com/covid19/social-distancing-scoreboard>.
14. Ferres, Leo, Rossano Schifanella, Nicola Perra, Salvatore Vilella, Loreto Bravo, Daniela Paolotti, Giancarlo Ruffo, and Manuel Sacasa. “Measuring Levels of Activity in a Changing City,” n.d., 11.
15. Huang, Xiao, Zhenlong Li, Yuqin Jiang, Xiaoming Li, and Dwayne Porter. “Twitter Reveals Human Mobility Dynamics during the COVID-19 Pandemic.” *PLOS ONE* 15, no. 11 (November 10, 2020): e0241957. <https://doi.org/10.1371/journal.pone.0241957>.
16. J. Zahuranec, Andrew, Stefaan Verhulst, Andrew Young, Aditi Ramesh, and Brennan Lake. “What Is Mobility Data? Where Is It Used?” *GovLab*, Cuebiq, April 2021. http://files.thegovlab.org/COVID-19_Mobility_Data_One-Page_Brief_v3_041521.pdf.
17. Lai, Shengjie, Nick W. Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R. Floyd, Amy Wesolowski, et al. “Effect of Non-Pharmaceutical Interventions to Contain COVID-19 in China.” *Nature* 585, no. 7825 (September 1, 2020): 410–13. <https://doi.org/10.1038/s41586-020-2293-x>.
18. “Mobility and COVID-19 - Teralytics.” Accessed January 31, 2022. <https://www.teralytics.net/mobility-and-covid-19/>.
19. Schlosser, Frank, Vedran Sekara, Dirk Brockmann, and Manuel Garcia-Herranz. “Biases in Human Mobility Data Impact Epidemic Modeling.” *ArXiv:2112.12521 [Physics, q-Bio]*, December 23, 2021. <http://arxiv.org/abs/2112.12521.21>.

20. Warner AT, Obrecht A. Standardising humanitarian data for a better response: The Humanitarian eXchange Language. London: ODI/ALNAP, 2016.
21. Tatem AJ, Qiu Y, Smith DL, Sabot O, Ali AS, Moonen B. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malar J* 2009; 8: 287.
22. Toole JL, Herrera-Yaque C, Schneider CM, González MC. Coupling human mobility and social ties. *J R Soc Interface* 2015; 12: 20141128.
23. Bengtsson L, Gaudart J, Lu X, et al. Using mobile phone data to predict the spatial spread of cholera. *Sci Rep* 2015; 5: 8923.
24. Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS Med* 2011; 8: e1001083.
25. Wesolowski A, Winter A, Tatem AJ, et al. Measles outbreak risk in Pakistan: exploring the potential of combining vaccination coverage and incidence data with novel data-streams to strengthen control. *Epidemiol Infect* 2018; 146: 1575–83.
26. Wesolowski A, Buckee CO, Bengtsson L, Wetter E, Lu X, Tatem AJ. Commentary: Containing the Ebola Outbreak - the Potential and Challenge of Mobile Network Data. *PLoS Curr* 2014; 6: ecurrents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e.
27. Wesolowski A, Metcalf CJE, Eagle N, et al. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. *PNAS* 2015; 112: 11114–9.
28. Wesolowski A, Eagle N, Tatem AJ, et al. Quantifying the Impact of Human Mobility on Malaria. *Science* 2012; 338: 267–70.
29. Wesolowski A, Qureshi T, Boni MF, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *PNAS* 2015; 112: 11887–92.
30. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJE. Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data. *The Journal of Infectious Diseases* 2016; 214: S414–20.
31. How Telenor works with Big Data for Social Good. Telenor Group. <https://www.telenor.com/sustainability/how-telenor-works-with-big-data-for-social-good/> (accessed Nov 21, 2021).
32. Rajput R. Maharashtra deploys drones, uses CDR, boosts patrolling to contain Covid-19. *The Economic Times*. 2020; published online April 6. <https://economictimes.indiatimes.com/news/politics-and-nation/maharashtra-deploys-drones-uses-cdr-boosts-patrolling-to-contain-covid-19/articleshow/74999989.cms?from=mdr> (accessed Nov 21, 2021).
34. Inferring Jakarta Commuting Statistics from Twitter • UN Global Pulse. 2017; published online May 30. <https://www.unglobalpulse.org/2017/05/inferring-jakarta-commuting-statistics-from-twitter/> (accessed Oct 23, 2021).
35. Tufekci Z. How social media took us from Tahrir Square to Donald Trump. *MIT Technology Review*. 2018; published online Aug 14. <https://www.technologyreview.com/2018/08/14/240325/how-social-media-took-us-from-tahrir-square-to-donald-trump/> (accessed Oct 23, 2021).
36. Chander A, Kaminski ME, McGeeveran W. Catalyzing Privacy Law. Rochester, NY: Social Science Research Network, 2019 DOI:10.2139/ssrn.3433922; https://web.archive.org/web/20211023132113/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3433922.
37. Camber Systems. Camber. <https://cambersystems.com/> (accessed Nov 21, 2021).
38. Cuebiq. Cuebiq. <https://www.cuebiq.com/> (accessed Nov 21, 2021).
39. Magicbox – Partnerships. <https://www.unicef.org/innovation/magicbox-partnerships> (accessed Nov 21, 2021).
40. CrisisReady. CrisisReady. <https://www.crisisready.io/> (accessed Nov 21, 2021).
41. Data for Good. <https://dataforgood.facebook.com/> (accessed Nov 21, 2021).



42. Pasternack SM and A. Here are the data brokers quietly buying and selling your personal information. Fast Company. 2019; published online March 2. <https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information> (accessed Nov 21, 2021).
43. Unacast. <https://www.unacast.com/> (accessed Nov 21, 2021).
44. "Data-Driven Development. Pathways for Progress." World Economic Forum, January 2015. https://www3.weforum.org/docs/WEFUSA_DataDrivenDevelopment_Report2015.pdf.
45. Verhulst S, Zahuranec AJ, Young A, Ramesh A. The Use of Mobility Data for Responding to the COVID-19 Pandemic. Rochester, NY: Social Science Research Network, 2021 <https://papers.ssrn.com/abstract=3937659> (accessed Nov 21, 2021).
46. Google COVID-19 Community Mobility Reports. <https://www.google.com/covid19/mobility/> (accessed Oct 23, 2021).
47. "Data-Driven Development. Pathways for Progress." World Economic Forum, January 2015. https://www3.weforum.org/docs/WEFUSA_DataDrivenDevelopment_Report2015.pdf.
48. Beyond "Do No Harm" and Individual Consent: Reckoning with the Emerging Ethical Challenges of Civil Society's Use of Data Nathaniel A. Raymond in Group Privacy: New Challenges of Data Technologies
49. Meta Research. "Protecting Privacy in Facebook Mobility Data during the COVID-19 Response - Meta Research." Accessed January 31, 2022. <https://research.facebook.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/>.
50. Humanitarian Data Exchange. <https://data.humdata.org/> (accessed May 25, 2018).
51. Mueller U. Common Types of Partnerships Related to ICTS and Data. 2020; : 6.
52. New NetHope Report: Facebook Data for Good population density maps and how they are helping the nonprofit sector - NetHope. <https://nethope.org/2021/07/09/new-nethope-report-facebook-data-for-good-population-density-maps-and-how-they-are-helping-the-nonprofit-sector/> (accessed Nov 21, 2021).
53. Chawla S, Dwork C, McSherry F, Smith A, Wee H. Toward Privacy in Public Databases. In: Kilian J, ed. Theory of Cryptography. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 363–85.
54. Dinur I, Nissim K. Revealing information while preserving privacy. In: Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '03. San Diego, California: ACM Press, 2003: 202–10.
55. Adam NR, Worthmann JC. Security-control methods for statistical databases: a comparative study. *ACM Comput Surv* 1989; 21: 515–56.
56. Garfinkel S, Abowd J, Martindale C. Understanding database reconstruction attacks on public data. *Communications of the ACM* 2019; 62: 46–53.
57. Costanza-Chock S. Design Justice, A.I., and Escape from the Matrix of Domination. *Journal of Design and Science* 2018; published online July 16. DOI:10.21428/96c8d426.
58. Hill K. The facial-recognition app Clearview sees a spike in use after Capitol attack. *The New York Times*. 2021; published online Jan 9. <https://www.nytimes.com/2021/01/09/technology/facial-recognition-clearview-capitol.html> (accessed Oct 23, 2021).
59. Hill K. The Secretive Company That Might End Privacy as We Know It. *The New York Times*. 2020; published online Jan 18. <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> (accessed Oct 23, 2021).
100. Chang HH, Wesolowski A, Sinha I, et al. Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. *eLife* 2019; 8: e43481.
101. Mahmud AS, Kabir MI, Engø-Monsen K, et al. Megacities as drivers of national outbreaks: The 2017 chikungunya outbreak in Dhaka, Bangladesh. *PLOS Neglected Tropical Diseases* 2021; 15: e0009106.

102. Kishore N, Taylor AR, Jacob PE, et al. Evaluating the reliability of mobility metrics from aggregated mobile phone data as proxies for SARS-CoV-2 transmission in the USA: a population-based study. *The Lancet Digital Health* 2021; published online Nov 2. DOI:10.1016/s2589-7500(21)00214-4.
103. Cowley LA, Afrad MH, Rahman SIA, et al. Genomics, social media and mobile phone data enable mapping of SARS-CoV-2 lineages to inform health policy in Bangladesh. *Nature Microbiology* 2021; 6: 1271–8.
104. Grantz KH, Meredith HR, Cummings DAT, et al. The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nat Commun* 2020; 11: 4961.
105. OCHA. "Anticipatory Action," May 10, 2021. <https://www.unocha.org/our-work/humanitarian-financing/anticipatory-action>.
106. Kraemer MUG, Yang C-H, Gutierrez B, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 2020; 368: 493–7.
107. Pepe E, Bajardi P, Gauvin L, et al. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Sci Data* 2020; 7: 230.
108. Gao S, Rao J, Kang Y, Liang Y, Kruse J. Mapping County-Level Mobility Pattern Changes in the United States in Response to COVID-19. Rochester, NY: Social Science Research Network, 2020 DOI:10.2139/ssrn.3570145; https://web.archive.org/web/20211121184338/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3570145.
109. Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* 2020; 20: 1247–54.
110. Data For Good at Meta Health Facility Travel Times and Service Areas. <https://dataforgood.facebook.com/dfg/tools/health-facility-travel-times-and-service-areas> (accessed Dec 8, 2021).
111. COVID-19 Mobility Data Network. COVID-19 Mobility Data Network. <https://www.covid19mobility.org/> (accessed Nov 21, 2021).
112. Nissenbaum H. Privacy As Contextual Integrity. *Washington Law Review* 2004; 79: 39.
113. World Development Report 2021 team. World Development Report 2021: Data for Better Lives. <https://wdr2021.worldbank.org/stories/governing-data/> (accessed Nov 21, 2021).
114. Ong P, Ong J, Ong E, Carrasquillo A. Neighborhood Inequality in Shelter-in-Place Burden: Impacts of COVID-19 in Los Angeles. 2020; published online April 22. <https://escholarship.org/uc/item/0d87b0t5> (accessed Oct 23, 2021).
115. The war inside Palantir: Data-mining firm's ties to ICE under attack by employees. *Washington Post*. <https://www.washingtonpost.com/business/2019/08/22/war-inside-palantir-data-mining-firms-ties-ice-under-attack-by-employees/> (accessed Oct 23, 2021).
116. Jakarta PL. Strengthening Humanitarian Action with Insights from Mobile Network Data. *Pulse Lab Jakarta*. 2019; published online Dec 19. <https://medium.com/pulse-lab-jakarta/strengthening-humanitarian-action-with-insights-from-mobile-network-data-a7d158d71cbc> (accessed Nov 21, 2021).
117. "Anticipatory Action | OCHA." Accessed February 10, 2022. <https://www.unocha.org/our-work/humanitarian-financing/anticipatory-action>.





Appendices

Action Points: Overview

Data Readiness

- 1. Introduce legislation to govern the use (and re-use) of such novel data streams, including human mobility data.*
- 2. Launch a professional body comprising technology companies, researchers, and response agencies to publish interoperability standards on human mobility data and other novel data streams.*
- 3. Generate context-specific consensus on aggregation and anonymization of shared data.*
- 4. Promote the development of standard contractual language for the use of human mobility data by academics and policy makers.*
- 5. Promote a cadre of data stewards within technology companies to guide the responsible and meaningful use of the data for public good.*
- 6. Develop, test, and disseminate use-cases on the application of differential privacy on human mobility data sets for epidemiological or other public health purposes.*
- 7. Include communities from whom the data are generated in defining the scope of use of data.*

Methods Readiness

- 1. Develop frameworks for communicating bias and uncertainty while publishing and communicating analyses using mobility data from mobile phones.*
- 2. Advance methods to address bias correction in human mobility data sets.*
- 3. Support the development of standards to promote interoperability among the data sets, especially across temporal and spatial scales.*
- 4. Develop approaches to allow for the diverse anonymization techniques used across technology companies.*
- 5. Develop a framework for quantifying the potential for harm that acknowledges societal and political context.*
- 6. Advocate for donors and national academies to greatly invest resources for translational data science, in preparation for and during emergencies, including allocating emergency funds.*

Translational Readiness

1. Document and disseminate best practices targeting:

- a. technology companies, to incentivize them to keep sharing these data responsibly in the absence of a regulatory mandate to do so;*
- b. donors and national academies, to incentivize them to greatly invest resources for translational data science in preparation for and during emergencies. Stronger investments in advancing both the methodologies as well as the local capacity to receive and act on the sights, are required to generate the evidence that the data are useful. During emergencies, provide funds to accelerate translational work.*
- c. policymakers, to socialize them to the potential use of these data products and to the expertise required to generate, process, and analyze them, and to help response agencies guide purposeful, actionable analytic products;*
- d. general public, to improve transparency about the potential use and limitations of these data to promote informed adoption.*

2. Fund and develop a cadre of “data bilinguals” embedded in public health and response agencies.

3. Fund and sustain a distributed network of researchers to support local regional collaborations of trained scientists and response agencies.

Seminar Goals

In 2020, we witnessed an acceleration in the availability and use of novel (big) data streams from mobile operations, AdTech, and social media companies to inform emergency response and planning. The use of these data to model public health scenarios and inform preparedness and response to natural disasters presents important methodological, ethical, and translational (capacity) questions.

Effective and equitable use of these data in emergency response calls for collaboration across three pillars: data readiness, methods readiness, and translational readiness. Accordingly, this seminar series convened an interdisciplinary group of experts from Industry, Academia, Law and Policy, and Society & Practice to identify recent advances and priorities ahead. The group sought to build consensus around key technical, ethical, and policy issues that must be addressed to facilitate the safe, responsible, fair, and equitable use of novel data streams to inform emergency preparedness and response.

The seminar series resulted in this white paper, which summarizes key points of consensus and recommendations. This document will in turn set the agenda for a consultative process (in collaboration with CrisisReady and the Global Facility for Disaster Reduction and Recovery, GFDRR) to develop guidance for governments and response agencies seeking to use these novel data streams for emergency preparedness and response.



Seminar Participants

Seminar I: Industry Participants

Rakesh Bharania, *Salesforce*
Maggie Carter, *Amazon Web Services (AWS)*
Olivier Cottray, *ESRI*
Alex Dow, *Meta*
Kenth Engø-Monsen, *Telenor*
Brennan Lake, *Cuebig*
Navin Vembar, *Camber Systems*
Andrew Zolli, *Planet*

Seminar I: Academia Participants

Mercè Crosas, *Institute for Quantitative Social Science (IQSS), Harvard University*
Marta Gonzalez, *UC Berkeley College of Environmental Design*
Sune Lehmann, *Technical University of Denmark*
Richard Maude, *Harvard T.H. Chan School of Public Health; Mahidol Oxford Tropical Medicine Research Unit; University of Oxford*
Elaine Okanyene Nsoesie, *Boston University*
Andy Tatem, *University of Southampton*
Salil Vadhan, *Harvard SEAS*
Li Xiong, *Emory University*

Seminar II: Law and Policy Participants

Jos Berens, *United Nations Office for the Coordinator of Humanitarian Affairs (OCHA)*
James De France, *International Federation of the Red Cross and Red Crescent Societies*
Urs Gasser, *TUM School of Social Sciences and Technology at the Technical University of Munich; Formerly Berkman Klein Center, Harvard University*
Laura Hecht-Felella, *Brennan Center for Justice, New York University*
Dragana Kaurin, *Berkman Klein Center, Harvard University*
Rahul Matthan, *Trilegal, India*
Laura McGorman, *Data for Good, Meta*
Aishwarya Raman, *Ola Mobility Institute*
Nathaniel Raymond, *Jackson Institute for Global Affairs, Yale University*
Emanuele Strano, *MindEarth*
Stefaan Verhulst, *The Governance Lab (GovLab), New York University*

Seminar II: Society and Practice Participants

Raquel Bernedo Pardal, *Spanish Red Cross*
Leremy Colf, *U.S. Department of Health and Human Services (HHS), Office of the Assistant Secretary for Preparedness and Response (ASPR)*
Lawrence Dimailig, *ASEAN Coordinating Centre for Humanitarian Assistance on Disaster Management (AHA Centre)*
Erin Hughey, *Pacific Disaster Center*
Daniel Joseph, *International Federation of the Red Cross and Red Crescent Societies*

Petrarca Karetji, *Pulse Lab Jakarta*
Keith Landicho, *ASEAN Coordinating Centre for Humanitarian Assistance on Disaster Management (AHA Centre)*
Derrick Munneene, *World Health Organization*
DJ Patil, *Devoted Health*
Linda Raftree, *Independent Consultant*
Vance Taylor, *California Governor's Office of Emergency Services (CalOES)*
Jason Vargo, *Federal Reserve Bank of San Francisco*

Organizing Team

Satchit Balsari, *Harvard Medical School; Beth Israel Lahey Health; Harvard T.H. Chan School of Public Health*
Abhishek Bhatia, *Harvard T.H. Chan School of Public Health*
Caroline Buckee, *Harvard T.H. Chan School of Public Health*
Jennifer Chan, *Northwestern University*
Maham Faisal Khan, *Global Facility for Disaster Reduction and Recovery (GFDRR), World Bank*
Nicholas Jones, *Global Facility for Disaster Reduction and Recovery (GFDRR), World Bank*
Nishant Kishore, *Harvard T.H. Chan School of Public Health*
Sraavya Sambara, *Harvard University*
Andrew Schroeder, *Direct Relief*
Navin Vembar, *Camber Systems*
Takahiro Yabe, *Global Facility for Disaster Reduction and Recovery (GFDRR), World Bank*





Visit CrisisReady Online

<https://crisisready.io>

Contact CrisisReady

info@crisisready.io

communication@crisisready.io

In Affiliation With

