The Use of Mobility Data for Responding to the COVID-19 Pandemic

DATA4COVID19
Deep Dive

ODI THEGOVLAB cuebiq

Acronyms and Abbreviations

CDR	Call Detail Records
COVID-19	Coronavirus disease 2019
CSR	Corporate Social Responsibility
GDPR	General Data Protection Regulation
GPS	Global Positioning System
ID	Identification
ODM	Origin-Destination Matrix
NYC	New York City
POI	Point of Interest
SDK	Software Development Kit
SMS	Short Message Service
Telcos	Telecommunications company
UN	United Nations
UNICEF	United Nations Children's Fund
US	United States
x-DR	x-Data Record

Table of Contents

1. Executive Summary 4
- 1.A. Methodology 7
- 1.B. Research Limitations 10

2. Mobility Data: What It Is and How It Can Be Used 11
- 2.A. Mobility Data Overview 12
- 2.B. Mobility Data for COVID-19 16
- 2.C. Users of Mobility Data 17
- 2.D. Approaches to Leveraging Mobility Data: Data Collaboration 18

3. Current Practice: Insights from Five Case Studies 22
- 3.A. Comparative Insights 22
- 3.B. Insights from Data Users 24
- 3.C. Insights from Data Suppliers 27

4. Prescriptive Analysis: Recommendations for the Future 30

5. About the Authors 33

6. Appendix 1: Mobility Data Type 35

7. Appendix 2: Data4Covid Mobility Repository: Taxonomy of Factors, Design Choices, and Variables 38

1. Executive Summary

As the COVID-19 pandemic continues to upend the way people move, work, and gather, governments, businesses, and public health researchers have looked increasingly at mobility data to support pandemic response. This data, assets that describe human location and movement, generally has been collected for purposes directly related to a company's business model, including optimizing the delivery of consumer services, supply chain management or targeting advertisements. However, these call detail records, smartphone-mobility data, vehicle-derived GPS, and other mobility data assets can also be used to study patterns of movement. These patterns of movement have, in turn, been used by organizations to forecast disease spread and inform decisions on how to best manage activity in certain locations.

Researchers at The GovLab and Cuebiq, supported by the Open Data Institute, identified 51 notable projects from around the globe launched by public sector and research organizations with companies that use mobility data for these purposes. It curated five projects among this listing that highlight the specific opportunities (and risks) presented by using this asset. Though few of these highlighted projects have provided public outputs that make assessing project success difficult, organizations interviewed considered mobility data to be a useful asset that enabled better public health surveillance, supported existing decision-making processes, or otherwise allowed groups to achieve their research goals.

The report below summarizes some of the major points identified in those case studies. While acknowledging that location data can be a highly sensitive data type that can facilitate surveillance or expose data subjects if used carelessly, it finds mobility data can support research and inform decisions when applied toward narrowly defined research questions through frameworks that acknowledge and proactively mitigate risk. These frameworks can vary based on the individual circumstances facing data users, suppliers, and subjects. However, there are a few conditions that can enable users and suppliers to promote publicly beneficial and responsible data use and overcome the serious obstacles facing them.

For data users (governments and research institutions), functional access to real-time and contextually relevant data can support research goals, even though a lack of data science competencies and both short and long-term funding sources represent major obstacles for this goal. Data suppliers (largely companies), meanwhile, need governance structures and mechanisms that facilitate responsible re-use, including data re-use agreements that define who, what, where, and when, and under what conditions data can be shared. A lack of regulatory clarity and the absence of universal governance and privacy standards have impeded effective and responsible dissemination of mobility for research and humanitarian purposes. Finally, for both data users and suppliers, we note that collaborative research networks that allow organizations to seek out and provide data can serve as enablers of project success by facilitating exchange of methods and resources, and closing the gap between research and practice.

Based on these findings, we recommend the development of clear governance and privacy frameworks, increased capacity building around data use within the public sector, and more regular convenings of ecosystem stakeholders (including the public and data subjects) to broaden collaborative networks. We also propose solutions towards making the responsible use of mobility data more sustainable for long-term impact beyond the current pandemic. A failure to develop regulatory and governance frameworks that can responsibly manage mobility data could lead to a regression to the ad hoc and uncoordinated approaches that previously defined mobility data applications. It could also lead to disparate standards about organizations' responsibilities to the public.

PHOTO: Unsplash/Dele Oke is licensed under CC0

With over 120 million cases and 2.6 million deaths as of March 2021, the COVID-19 pandemic continues to take a catastrophic toll.[1] With health care systems across the world under pressure, governments continue to limit the free movement of people to control virus spread. Meanwhile, people continue to face economic and social hardship from the pandemic. In some regions that successfully controlled the pandemic, additional waves of spread are forcing populations back into lockdown.

In addition to a relatively high fatality rate, COVID-19 is also easy to spread, with many of those infected spreading it to one or more people they know.[2] These factors make understanding people's interactions of the utmost importance. Traditionally, understanding connections between groups has been done by public health professionals through modeling based on existing epidemiological, population, and statistically derived travel data.[3] Given the unique challenges presented by COVID-19, however, some governments and researchers have looked toward new data sources to supplement existing approaches.

One of these sources is mobility data, data about the location of a device passively produced through normal activity. Mobility data generally has been collected for purposes directly related to a company's business model, including optimizing the delivery of consumer services or targeting advertisements. However, some professionals have used call detail records, smartphone- and vehicle-derived data, and

1. The Center for Systems Science and Engineering at Johns Hopkins University (JHU). "COVID-19 Map." Johns Hopkins Coronavirus Resource Center. Accessed January 5, 2021. https://coronavirus.jhu.edu/map.html.
2. Beech, Peter. "What Is COVID-19's R Number – and Why Does It Matter?" World Economic Forum, May 8, 2020. https://www.weforum.org/agenda/2020/05/covid-19-what-is-the-r-number/; Worldometer. "Coronavirus Death Rate (COVID-19) - Worldometer," May 14, 2020.
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6227252/

other assets to understand in near-real time where and when disease spread is likely to occur.[4] These insights can, in turn, inform official policies and help people make decisions about the individual contexts in which they operate.

Mobility insights and decision-support tools generated from de-identified and aggregated telecommunications and mobility data have already been used in response to public health threats such as Ebola in Africa,[5] Zika in Brazil[6] and swine flu in Mexico.[7] The use of location data in each of these cases were broadly considered successful in addressing serious and immediate public needs. However, uses of location data have also been used for more questionable purposes. An investigation in the United States found one health smartphone app developed for North and South Dakota shared user data alongside “advertising identifiers” with other companies, despite the app’s privacy policy saying no such sharing would occur.[8] In South Korea, poor security practices in a government-run health app made the names and real-time locations of quarantined individuals retrievable by hackers.[9] In China, the New York Times reported that the government’s Alipay app secretly shared the user’s location and other information with police.[10] In other contexts, companies that collect location data have reportedly sold their assets to law enforcement, military, and intelligence services without the knowledge or consent of users.[11]

Incompetence and malice around the use of location data is a very real threat, as is the continued failure to hold bad actors accountable for irresponsible practices. With governments around the world currently engaged in large-scale efforts to use mobility data insights to help them manage the current COVID-19 crisis, it is crucial to encourage good practices and to foster regulatory clarity and institutional safeguards. Responsible, sustainable, and effective public health surveillance is only possible if organizations commit themselves to protecting the rights and well-being of data subjects.

This report looks at several attempts by decision-makers in the public and private sector to engage with one another to address the ongoing crisis to see what promising practices exist. Many of these efforts have taken the form of data collaboratives, a new form of public-private partnership in which participants from different sectors exchange or provide access to expertise and data to create public value.[12] By mapping and studying these mobility data collaboratives, we seek to identify what makes a successful initiative and what does not. This work can be used not only to inform ongoing response and recovery efforts but to improve how institutions operate in a post-pandemic environment. It can support work to make data collaboration as a whole more responsible, sustainable, and systematic. From this basis, we approach the analysis by asking:

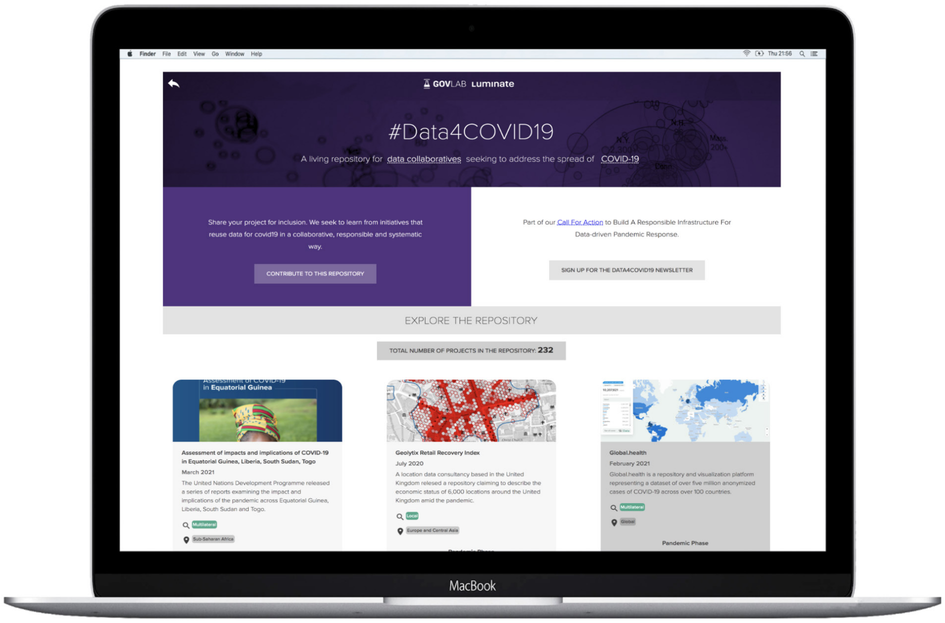
- What are possible conditions for success and failure of mobility data sharing initiatives within the context of COVID-19?

---

4. Yates, Christian. “How to Model a Pandemic.” The Conversation. Accessed March 15, 2021. http://theconversation.com/how-to-model-a-pandemic-134187.
5. Wesolowski, Amy, Caroline O. Buckee, Linus Bengtsson, Erik Wetter, Xin Lu, and Andrew J. Tatem. “Commentary: Containing the Ebola Outbreak - the Potential and Challenge of Mobile Network Data.” *PLoS Currents* 6 (September 29, 2014). https://doi.org/10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e.
6. Peiró, Patricia. “Los billetes de avión predicen dónde viajan las enfermedades.” *El País*. November 1, 2018, sec. Planeta Futuro. https://elpais.com/elpais/2018/10/29/planeta_futuro/1540825274_349206.html.
7. Oliver, Nuria. “Combating Global Epidemics with Big Mobile Data.” the Guardian, September 5, 2013. http://www.theguardian.com/media-network/media-network-blog/2013/sep/05/combating-epidemics-big-mobile-data.
8. Fowler, Geoffrey A. “One of the First Contact-Tracing Apps Violates Its Own Privacy Policy.” *Washington Post*, May 21, 2020, sec. Perspective. https://www.washingtonpost.com/technology/2020/05/21/care19-dakota-privacy-coronavirus/.
9. Sang-Hun, Choe, Aaron Krolik, Raymond Zhong, and Natasha Singer. “Major Security Flaws Found in South Korea Quarantine App.” *The New York Times*, July 21, 2020, sec. Technology. https://www.nytimes.com/2020/07/21/technology/korea-coronavirus-app-security.html.
10. Mozur, Paul, Raymond Zhong, and Aaron Krolik. “In Coronavirus Fight, China Gives Citizens a Color Code, With Red Flags.” *The New York Times*, March 2, 2020, sec. Business. https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html.
11. Cox, Joseph. “Military Unit That Conducts Drone Strikes Bought Location Data From Ordinary Apps,” March 4, 2021. https://www.vice.com/en/article/y3g97x/location-data-apps-drone-strikes-iowa-national-guard.
12. Data Collaboratives. “Data Collaboratives Home Page,” 2018. http://datacollaboratives.org/explorer.html.

- What are the barriers and challenges to successful implementation and outcomes? What are the risks involved? Are they unique to mobility data or indicative of data sharing partnerships in general?
- What tangible effects or impacts have there been from sharing, or not sharing, between the public and private sectors?
- What decision-making has been made possible from these sectors successfully sharing their mobility data?
- What are the prospects for sustained mobility data sharing programs in a post COVID-19 world?

1.A. Methodology



To answer these research questions, the Open Data Institute supported a research team composed of researchers from The GovLab and Cuebiq to assemble a repository of mobility data collaboratives related to COVID-19 and describe key attributes of each initiative. This mobility-focused repository stemmed from The GovLab's ongoing curation and review of data collaboratives for COVID-19 through its the #Data4COVID19 Living Repository.[13] Researchers sourced this sample of mobility data collaboratives through an analysis of this repository as well as additional desk research and expert consultations.

While The GovLab assembled most of the examples through desk research, subject-matter experts in mobility data from Cuebiq captured key attributes of these data collaboratives—including the type of mobility data shared and technical means for ensuring responsible data use. Once the research team compiled the repository, it selected five data collaboratives in coordination with the Open Data Institute to study further. These selections sought, to the extent possible, to explore projects across different regions that used different data collaborative types, project scopes, and data types. The team did, however, decide only to pursue those projects where substantial supporting documentation was available and where it would be possible to secure interviews with key stakeholders.

13. Originally launched in March 2019, the crowdsourced repository features, at the time of writing, over 200 global examples of data collaboratives initiated to address the COVID-19 pandemic. *See* http://list.data4covid19.org

Projects by Data Collaborative Type	
Intelligence Generation	12
Research and Analysis Partnership	20
Public Interface	9
Trusted Intermediary	3
Data Pooling	5
Prizes & Challenges	2
Total	**51**

Projects by Region	
East Asia and Pacific	2
Global	10
Europe and Central Asia	18
Latin America and the Caribbean	2
North America	18
Sub-Saharan Africa	1
Total	**51**

Projects by Sector	
Public Health	31
Social	16
Economic	4
Total	**51**

Projects by Insight	
Situational Awareness	29
Impact Assessment	14
Cause and Effect Analysis	15
Prediction	5
Total	**51**

By synthesizing insights gleaned from the detailed case studies and the broader repository, we provide a view into the current state of mobility data collaboratives for COVID-19 response while also identifying conditions for success and obstacles to achieving impact. Based on this diagnosis, we provide recommendations towards enabling more systematic, sustainable and responsible models of reuse of mobility data through data collaboration as to support decision-making regarding pandemic prevention, monitoring, and response.[14] This document summarizes the results of that work. It includes several sections:

- **Section II** of this document provides an introduction to mobility data, the value and characteristics of this data as it relates to COVID-19, and the organization of our Mobility Data for COVID-19 repository. It also includes a brief explanation of data collaboration, the structure through which the data described in this document is shared.
- **In Section III,** we provide a diagnostic analysis of the contents of the Mobility Data for COVID-19 repository and the five case studies that were subject to in-depth analysis. These projects include:
    - Facebook Data for Good;

14. Data Collaboratives. "Data Collaboratives Home Page," 2018. http://datacollaboratives.org/explorer.html.

- The European Commission Joint Research Centre's Mobility Functional Area Study;
- The Greater London Authority Data Store's COVID-19 Mobility Report;
- Telefónica Chile and Universidad del Desarrollo Mobility Reports; and
- The University of Toronto's Canadian Mobility Analysis.

We detail insights that can be gleaned from these insights for demand-side actors, those who need data for analysis and supply-side actors, those who provide data.

- **Section IV** provides a prescriptive analysis, including an overview of enabling factors and preconditions of success. It also includes recommendations for improving current practice regarding COVID-19 and a set of recommendations for the field of data. collaboration more generally.[15]

1.B. Research Limitations

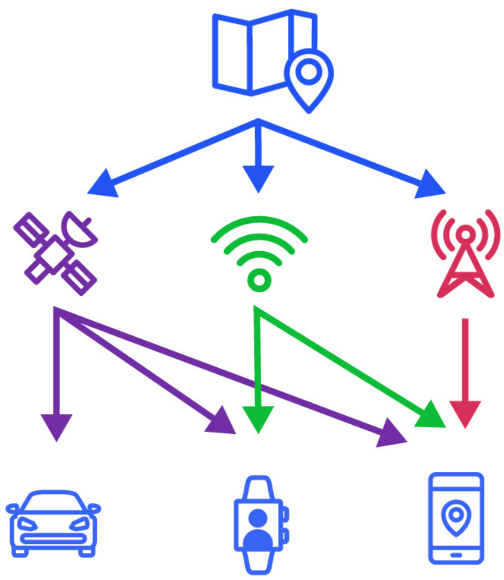
The following report, associated case studies, and repository rely on desk research and key stakeholder interviews. This approach represents a few limitations that readers should keep in mind. First, the reliance on public reporting means that the projects highlighted here are not comprehensive and, in some way, self-selected. The COVID-19 pandemic has generated a large number of data-driven initiatives, some of them well-known and others secretive. Both public and private sector entities have an incentive only to share information about projects that they consider clearly successful, meaning many projects with negative or ambiguous results are simply never reported by those who create them. Second, the frequent lack of tangible outputs, transparent processes, or third party involvement means that, in many of the examples highlighted, the team is reliant on how project sponsors characterize their own work and the work of their partners. While the research team sought documentation to validate these characterizations wherever possible and sought out multiple interviewees to get multiple perspectives on the assembled projects, researchers recognize the inherent limitations here.

The research team also recognizes the potential conflict of interest represented by the involvement of researchers from Cuebiq, a provider of mobility data and a participant in several of the projects listed. To promote research independence, the research team explicitly excluded Cuebiq staff from drafting and editing case studies on any projects Cuebiq was directly involved in. Recognizing the possibility of implicit bias—Cuebiq's involvement shaping the team's characterization of Cuebiq and its competitors—the reports clearly indicate Cuebiq's involvement in the overall effort wherever appropriate.

15. We acknowledge that the insights and recommendations detailed in section III and IV may not be applicable across all contexts, especially those in which mobile phone or other advanced data collection strategies may not be possible e.g. conflict zones, countries with low human development indices, etc.

2. Mobility Data: What It Is and How It Can Be Used

For the past decade researchers have used location data derived from mobile phones to analyze human mobility dynamics. The insights generated from mobility data informs many topics, including transportation and urban planning, emergency management, and disease surveillance.[16] Such data is collected by private sector entities such as telecommunication companies (telcos) through Call Detail Records (CDRs) or through smartphone applications that use mobile phones' built-in GPS capabilities. While private sector mobility data is primarily monetized through commercial activities related to advertising and analytics, some location data companies and telcos also provide access to mobility datasets for academic research and humanitarian purposes.



As with many types of emerging technologies, the use of location data in the private and academic realms serves as testing grounds for concepts that can later be adopted for public use. One notable example of this is the US Department of Transportation's use of Google's mobility analytics for improving public transportation, which helps cities understand "where citizens want to travel and how to transport them to those destinations more efficiently, fairly, and safely."[17] Another example is the United Nations' use of CDRs from Turk Telekom to assess social integration of refugees in Turkey; the UN created a

16. Frias-Martinez, E., G. Williamson, and V. Frias-Martinez. "An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information." In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, 57–64, 2011. https://doi.org/10.1109/PASSAT/SocialCom.2011.142.

Wang, Feilong, Xiangyang Guan, and Cynthia Chen. "Assessing Impacts of Abnormal Events on Travel Patterns Leveraging Passively Collected Trajectory Data." *ArXiv:1911.11633 [Physics]*, March 29, 2020. http://arxiv.org/abs/1911.11633.

Wang, H., F. Calabrese, G. Di Lorenzo, and C. Ratti. "Transportation Mode Inference from Anonymized and Aggregated Mobile Phone Call Detail Records." In *13th International IEEE Conference on Intelligent Transportation Systems*, 318–23, 2010. https://doi.org/10.1109/ITSC.2010.5625188.

17. TechCrunch. "US DOT and Alphabet's Sidewalk Labs Aim to Create New Public Transit and Wi-Fi Networks." Accessed January 5, 2021. https://social.techcrunch.com/2016/03/17/u-s-dot-and-sidewalk-labs-hustle-and-flow/.

methodological framework which allowed public sector organizations to assess the impacts of relief programs and the social, economic, and cultural burden placed on refugees.[18]

While mobility data holds the potential to create positive social impact through cases such as these, it can also be used for ethically problematic purposes. For instance, in early 2020, some reports alleged that one Muslim prayer app collected location data that fed into counterterrorism operations by the US military.[19] In countries with social and political unrest, the mis-use of mobility data can endanger citizens or asylum-seekers, for whom information about their whereabouts is particularly sensitive. Recent research has shown that location data is a key part of the extension of the "surveillance state."[20] Given the privacy sensitivities inherent in the mass collection of location data, it is critical to evaluate the mechanisms by which data is collected and shared in order to weigh utility against (group) privacy.

In addition to privacy concerns, it should also be noted that mobility data alone is not a panacea, and certain types of mobility data may under-represent segments of society due to the collection method. For example, mobility data derived from smartphones may under-represent those with lower rates of smartphone ownership, such as elderly populations and lower-income groups, especially in countries with overall lower smartphone penetration.[21]

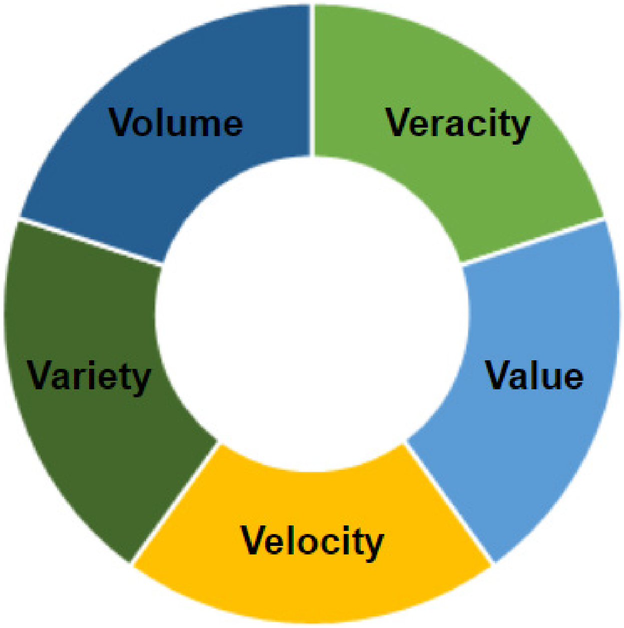
When COVID-19 emerged as a global health threat in early 2020, epidemiologists across research institutions and the public sector alike (in areas where resources for such analyses exist) turned to mobility data as a source for situational awareness to inform containment strategies. For instance, countries such as South Korea used smartphone-derived location data for surveillance activities related to contact tracing while companies such as Google and Apple developed technologies for bluetooth-enabled contact tracing.[22]

2.A. Mobility Data Overview

There are many mobility data sources and formats, each relying on different technologies and collection mechanisms. The type of location data used for COVID-19 mobility modeling is often derived from private sector sources such as telecommunications and smartphone app developers. However, it can also be sourced from public sector sources, such as public transportation authorities who can provide sensor or ridership data. Data provenance is an important factor that affects all "Five V's" of big data, detailed here as they relate to mobility data:

- **Volume:** the scale of data being collected both in terms of number of users and records per user;
- **Veracity:** the spatial precision and accuracy of data being collected;
- **Value:** the relevance of the data to the use case at hand;
- **Velocity:** the rate at which data is collected and delivered to the user; and
- **Variety:** the diversity of metadata being collected.

18. Pulse Lab New York. "Using Call Detail Records To Understand Refugee Integration In Turkey • UN Global Pulse." *UN Global Pulse* (blog), 2018. https://www.unglobalpulse.org/project/using-call-detail-records-to-understand-refugee-integration-in-turkey-2/.
19. Cox, Joseph. "How the U.S. Military Buys Location Data from Ordinary Apps." November 2020. https://www.vice.com/en/article/jgqm5x/us-military-location-data-xmode-locate-x.
20. Frith, Jordan and Michael Saker. "It Is All About Location: Smartphones and Tracking the Spread of COVID-19." July 30, 2020. https://doi.org/10.1177/2056305120948257
21. Pew Research Center. "Mobile Fact Sheet." *Internet & Technology* (blog), June 12, 2019. https://www.pewresearch.org/internet/fact-sheet/mobile/.
22. Hsu, Jeremy. "Contact Tracing Apps Struggle to Be Both Effective and Private - IEEE Spectrum." IEEE Spectrum, September 24, 2020. https://spectrum.ieee.org/biomedical/devices/contact-tracing-apps-struggle-to-be-both-effective-and-private.



In addition to the "five V's" are the ethical implications of both how the data is collected and how it is shared. Did the data collector obtain the informed consent of the user? If so, does the data collector enable users to exercise their data rights in accordance with regulations such as the European Union's General Data Protection Regulation (GDPR)? How does the data controller safeguard the privacy of individual users? These ethical concerns are of paramount importance, especially in countries where regulations around appropriate use often do not exist.

Below, we seek to provide clarity on all these issues. We provide a high-level overview of the types of mobility data that are used for COVID-19 analyses. (A more detailed technical description of each source can be found in the Appendix.) We also provide descriptions and terminology of data provenance in terms of ethical considerations—such as consent—and technical measures taken to preserve privacy.

An Emerging Taxonomy of Mobility Data

Data on mobility can be collected through various methods. Several notable sources of mobility data include:

1. **Call Detail Records (CDR):** Passively collected records of each time a mobile phone connects to a network by sending or receiving a voice call or SMS.[23] CDR generally benefits from a large scale of users, albeit with less accuracy than GPS data.

2. **x-Data Record (x-DR):** Any type of data created when a mobile phone user connects to a mobile network.[24] Like CDR, x-DR generally benefits from a large scale of users, albeit with less accuracy than GPS data.

3. **First-Party SDK smartphone-derived GPS:** Data transmitted on the geographic location of a smartphone using the device's built-in sensor. "First Party" implies the data holder collects data

23. Jones, Kerina Helen, Helen Daniels, Sharon Heys, and David Vincent Ford. "Public Views on Using Mobile Phone Call Detail Records in Health Research: Qualitative Study." *JMIR MHealth and UHealth* 7, no. 1 (January 16, 2019). https://doi.org/10.2196/11730.
24. Blackburn, Mandy. "Guide to Useful Telecoms Risk Assurance Acronyms and Terms." Risk Assurance Group, 2017. https://riskandassurancegroup.org/wp-content/uploads/2017/01/Generic-Guide-to-useful-Telecoms-Acronyms.pdf.

directly from individual users through a single collection methodology.

4. **Third-Party SDK smartphone-derived GPS:** Data transmitted on the geographic location of a smartphone by an application using the device's built-in sensor. "Third party" implies that the data holder obtains data from various third-party sources, and therefore lacks a direct relationship with users and requires data cleansing.

5. **BidStream Data:** Data about a potential recipient of a digital advertisement, including device identifiers and cookies, location data, IP address, and unique demographic and biometric information, by the organizations bidding to place an ad.[25] BidStream data is less precise than other sources. It also does not allow for longitudinal analyses, and presents ethical concerns since users have no way of opting in or out from data collection.

6. **Wearable GPS:** Data derived from wearable sensors—connected to the internet of things—that log the user's physical location.

7. **Aerial:** Aerial mobility data refers to data collected about physical movements of people based on images derived from aerial photography, often taken from satellites.[26]

8. **Vehicle GPS:** Data transmitted on the geographic location of a vehicle—such as an automobile, ship, or airplane—through in-vehicle sensors or other devices.

9. **Bluetooth:** A signal that allows for short-range wireless communication between electronic devices that can be used to determine when objects come into proximity with each other.[27]

10. **Geotagged social media data:** Metadata connected to videos, photographs, messages, and other content posted on social media containing information on the time and geographic location of the user.[28]

Consent path

Consent paths refer to the extent to which users of different technologies can consent to data collection. Governments, companies, and other stakeholders have control and responsibility for setting these practices.

- **Opt-in/Opt-out:** The concept of "opt-in" refers to an application, platform, or tool that allows users to consent to data collection. Consent is defined as "freely given, specific, informed and unambiguous" given by "clear and affirmative action."[29] The concept of "opt-out" refers to the ability of a user to exercise their rights to discontinue collection of their data. In addition to requiring that data collectors obtain consent to collect data from users, the European Union's GDPR gives users the right to "portability" and "erasure." Portability allows for users to request copies of their data while erasure, "the right to be forgotten," can allow users to demand that their data be permanently deleted.[30]

25. Wyden, Ron, Maria Cantwell, Elizabeth Warren, Anna Eshoo, Yvette Clarke, Bill Cassidy, Sherrod Brown, Edward Markey, Zoe Lofgren, and Ro Khanna. "Letter to the Honorable Joseph J. Simons, Chairman of the Federal Trade Commission," July 31, 2020. https://www.wyden.senate.gov/imo/media/doc/073120%20Wyden%20Cassidy%20Led%20FTC%20Investigation%20letter.pdf.
26. US Geological Survey. "National Geospatial Program." Accessed January 5, 2021. https://www.usgs.gov/core-science-systems/national-geospatial-program/imagery.
27. Figliola, Patricia Moloney. "Digital Contact Tracing Technology: Overview and Considerations for Implementation." Washington, DC: Congressional Research Service, May 29, 2020. https://www.everycrsreport.com/files/2020-05-29_IF11559_f93c95d-9673977192f24c0c149cbe28ca7f20e50.pdf.
28. https://www.sciencedirect.com/science/article/abs/pii/S0261517716301005?casa_token=tLhh5zlgjhkAAAAA:tW51e5tF-juT-plhzBxkuHsw2-1h79w-lFp_IHBBDh4cN_vaM5CGVd7zx2-uEr8tlnGLnz-c
29. Heimes. "How Opt-in Consent Really Works," February 22, 2019. https://iapp.org/news/a/yes-how-opt-in-consent-really-works/.
30. Individuals subject to the GDPR and similar laws have the right to request to be forgotten. However, this right of individuals

- **Opt-out only:** "Opt-out only" refers to an application, platform, or tool that collects data from users by default without first obtaining consent. Such policies only allow users to opt-out if they manually navigate to the system's settings and manually withdraw consent for such practices.
- **No explicit opt-in or opt-out:** A platform with no explicit opt-in or opt-out procedures is one that intentionally or unintentionally hides whether the app is collecting data on the user. Most jurisdictions require organizations to inform users whether their data is being collected.

Technical privacy measures

Since human mobility data is derived from inferring the locations and mobility of individual people, there are privacy issues inherent in using mobility data, no matter the intention of the analysis. Even when a mobility data has been provided with the informed consent of a user who also has full control over their data rights, data controllers must apply technical measures to the dataset to preserve the privacy of underlying users. The following examples demonstrate a range of techniques to preserve privacy, with varying degrees of privacy protection versus data utility.

De-identified: Many mobility data sources are collected from individual users with a unique identifier, such as a customer ID for CDRs or a mobile advertising ID for SDK-derived GPS data. While anonymizing these unique identifiers is an important step towards protecting privacy, anonymization alone has been proven to be an insufficient measure, since downstream users can attempt to re-identify individual users by combining multiple data sources, such as public home address records.[31] Since anonymized records alone are not necessarily aggregated, it may require significant data science and computational resources to process unaggregated anonymized mobility data.

De-identified with differential privacy: To preserve the utility of granular location data while taking a more robust approach than simple anonymization, data collectors can also apply various differential privacy techniques to individual data records, such as adding "noise" by modifying the latitude and longitude points and coarsening the level of accuracy within location data records. Additional measures include regularly shuffling anonymized IDs, removing sensitive points of interest from datasets, and obfuscating identifiable locations such as home locations to preclude attempts at re-identification.[32]

Aggregated: One of the most straightforward and effective methods for preserving privacy is through aggregation—the combining of multiple location data records into high level statistics.[33] Within the realm of mobility data for COVID-19, aggregation is often used to measure trends such as median distance traveled by all users within a given area, such as a county. In addition to preserving privacy, aggregated data has the added benefit in some instances of more easily fitting into pre-built epidemiological models. One potential downside is that once data is aggregated, it can be difficult to perform more fine-grained analyses.

On-premise federated: To avoid trade-offs between privacy preservation and the utility of granular analyses, data providers can offer sandbox environments that allow users to submit queries to datasets that remain on premise with the data provider. Upon submitting queries to the data provider, the user

to request removal of their data may not apply to large and anonymized databases where personal identifiers have been already removed. In December 2018, the Austrian Data Protection Authority ruled that anonymizing personal information satisfied the erasure requirement so long as neither the data controller nor a third party can "restore a personal reference without disproportionate effort." See: Marko, Roland, and Iris Riepan. "Austrian GDPR-Tracker: Data Protection Authority On Erasure By Way Of Anonymization - Privacy - Austria." Mondaq, February 11, 2019. https://www.mondaq.com/austria/data-protection/780272/austrian-gdpr-tracker-data-protection-authority-on-erasure-by-way-of-anonymization.

31. Montjoye, Yves-Alexandre de, Sébastien Gambs, Vincent Blondel, Geoffrey Canright, Nicolas de Cordes, Sébastien Deletaille, Kenth Engø-Monsen, et al. "On the Privacy-Conscientious Use of Mobile Phone Data." *Scientific Data* 5, no. 1 (December 11, 2018): 180286. https://doi.org/10.1038/sdata.2018.286.

32. ElSalamouny, Ehab, and Sebastien Gambs. "Differential Privacy Models for Location- Based Services," 2016, 34.

33. Kissner, Lea. "Aggregating over Anonymized Data," June 25, 2019. https://iapp.org/news/a/aggregating-over-anonymized-data/.

receives aggregated datasets in return. While users of such platforms require a higher degree of proficiency in data science, such a model provides the privacy preservation of aggregated data with a greater degree of customization and flexibility.[34]

Synthetic Mobility data: Another approach to benefit from the use of granular location data while protecting privacy is to develop synthetic mobility datasets. A synthetic data is identical in structure to a real-world dataset, but it is generated by mimicking the statistical distributions of a real dataset, without being linked to any individual user-generated data. One drawback of synthetic datasets for mobility is that it requires sophisticated machine learning technology to generate usable datasets.[35]

2.B. Mobility Data for COVID-19

Just as there are many types of data sources and formats, so too are there many ways mobility data can be used. For the purpose of this report, we examine mobility data collaboratives developed in response to the pandemic that focus on using data for gaining population-level insights into evolving mobility dynamics. Within this realm, mobility data has been used for a variety of analyses to ask questions and glean metrics including, for instance:

- **Average distance traveled by area:** How much are people moving and traveling now as compared to pre-pandemic times?
- **Origin-Destination between places:** What places do people travel to and from, and how can we use this information to model disease spread?
- **Proximity networks and social mixing:** How often are people coming into close proximity with each other, and what does this mean for transmissibility?
- **Point of interest visitation:** What types of places are people visiting more or less, and what impact does this have on the economy as well as potential for disease spread?
- **Mass migration and relocation patterns:** Are people leaving cities for more suburban and rural areas?
- **Quarantine adherence:** What percentage of the population is staying home during lockdowns?
- **Mobility inequality:** How might mobility and quarantine patterns differ between high and low income segments?
- **Spread forecasting:** How can we use all of these variables to forecast the spread of infectious disease?[36]

Through our research, alone, we developed a living repository representing a diverse array of data collaboratives featuring:

- 51 Unique Data Collaboratives
- 36 Unique Data Providers
- 136 Countries

34. Montjoye, Yves-Alexandre de, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. “OpenPDS: Protecting the Privacy of Metadata through SafeAnswers.” Edited by Tobias Preis. *PLoS ONE* 9, no. 7 (July 9, 2014): e98790. https://doi.org/10.1371/journal.pone.0098790.
35. El Emam, Khaled. “Accelerating AI with Synthetic Data.” IAPP. *Privacy Tech* (blog), February 26, 2020. https://iapp.org/news/a/accelerating-ai-with-synthetic-data/.
36. Oliver et al, *supra* note 7.

2.C. Users of Mobility Data

While the composition of each mobility data collaborative is unique, there are common trends around the needs of various data users and the type of mobility data provided by different data suppliers. As developed by the Open Data Institute, data users on the demand side can be categorized into four general user personae.[37] While any type of data user can unintentionally misuse data, these user types below do not include individuals seeking to intentionally pursue inappropriate uses of mobility data.

Researchers: Whether working within government agencies, the humanitarian sector, universities or private research institutions, researchers using mobility data typically have advanced expertise in data science, and are often motivated to develop novel approaches to working with data or applications for using mobility data. Researchers often require rich, granular datasets, and also benefit from historical data to perform longitudinal studies. An example of a “researcher” user persona can be found in our repository, and detailed case study on the COVID-19 Mobility Data Network, in which researchers from Harvard University and other institutions used mobility data from Facebook and Camber Systems for a variety of COVID-19 research topics.[38]

Local government workers: Local government workers are at the forefront of COVID-19 response within their local communities. Whether they work in small towns or large metropolitan areas, these users require mobility data that is localized and relevant to their communities. Local government workers also require more aggregated, user-friendly datasets, especially within smaller communities, which may not have data scientists as part of their staff. An example of local government workers leveraging mobility data that can be found in our repository is New York City’s Recovery Data Partnership, in which local officials use aggregated mobility data from Safegraph, Foursquare, and Cuebiq to inform local decision making related to COVID-19.[39] (Editor’s Note: Cuebiq is an author of this report.)

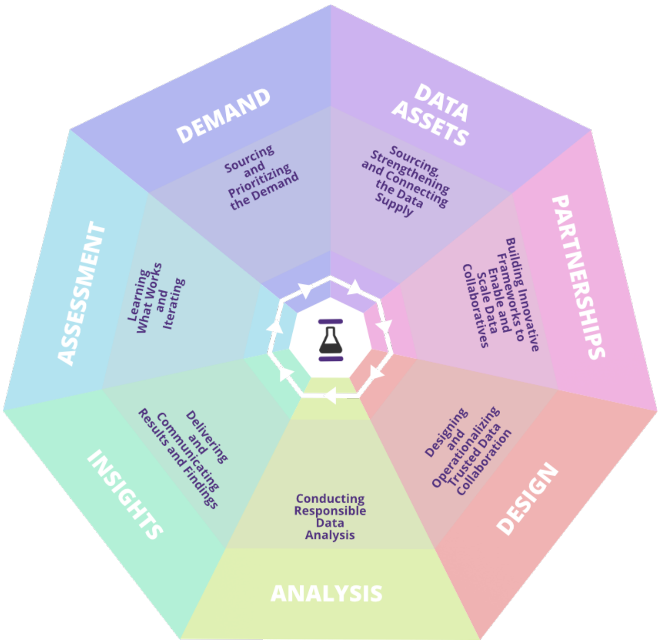
Public health policy officials: To create effective policies aimed at curbing the spread of COVID-19 and flattening the curve of caseloads, public health policy officials require situational awareness into macro trends of how people are moving between places, such as counties or cities. These users generally require aggregated datasets that can be readily standardized and incorporated into existing models, such as epidemiological models. An example public health policy officials using mobility data in this manner can be found within the Informes de Movilidad data collaborative in Chile, in which data from Telefonica Chile was aggregated into nation-wide mobility reports used by various agencies within the government of Chile.

Journalists: With the primary objective of informing the public, journalists use mobility data to provide easily digestible information on mobility trends through data journalism reporting that often includes statistics and interactive visualizations. Journalists tend to be generalists and therefore require user-friendly and generalizable datasets that have the potential to tell a story that is easily understood. Data journalists also benefit from frequently updated datasets so that they can continue to inform the public as the nature of the pandemic and its effects continue to evolve.

37. Knowledge & Opinion. “Covid-19 Mobility Data User Personas – The ODI.” Open Data Institute, September 28, 2020. https://theodi.org/article/covid-19-mobility-data-user-personas/.
38. COVID-19 Mobility Data Network. “COVID-19 Mobility Data Network.” COVID-19 Mobility Data Network, 2020. https://www.covid19mobility.org/.
39. NYC Mayor’s Office of Data Analytics. “Recovery Data Partnership.” NYC.gov, 2020. https://www1.nyc.gov/site/analytics/initiatives/recovery-data-partnership.page.

2.D. Approaches to Leveraging Mobility Data: Data Collaboration

What Are Data Collaboratives?



Data collaboratives are an emerging form of collaboration in which data or data expertise held by an entity in the private sector is leveraged in partnership with another entity (from the public sector, civil society, and/or academia) for public good. Data collaboratives draw together otherwise siloed data and a dispersed range of expertise, matching supply and demand and ensuring that relevant institutions and individuals are using and analyzing data in ways that maximize the possibility of new, innovative social solutions.[40]

Data collaboratives can be structured in many different ways but generally tend to involve one or more data holders from the private sector who can provide data (supply-side actors) and one or more actors working in the public sector, civil society, or academia that seek to use data in the public interest (demand-side actors). When designed well, supply- and demand-side actors come together based on a clear problem definition or organizing question; determine a fit-for-purpose operational model for providing functional access to data; enshrine a governance framework and any necessary legal documentation to organize decision-making, ensure accountability, and mitigate risks; and co-create a tactical roadmap for collaboratively moving from data to insight to action.[41]

40. Verhulst, Stefaan, Andrew Young, Michelle Winowatan, and Andrew J. Zahuranec. “Leveraging Private Data for Public Good: A Descriptive Analysis and Typology of Existing Practices.” Brooklyn, New York: The GovLab, October 2019. http://datacollaboratives.org/existing-practices.html.
41. Verhulst, Stefaan, and Andrew J. Zahuranec. “Using Data for COVID-19 Requires New and Innovative Governance Approaches.” *Data & Policy* (blog), May 21, 2020. https://medium.com/data-policy/using-data-for-covid-19-requires-new-and-innovative-governance-approaches-6c1350d9a2f2.

Typology of Data Collaboration

The GovLab introduced a typology of data collaboratives in 2019 informed by an analysis of over 100 projects from around the world included in the Data Collaboratives Explorer.[42] Two key variables formed the basis of the typology: 1) the level of engagement between data supply and data demand actors; and 2) the accessibility or level of conditionality of accessing data by external parties.

	Open Access	Restricted Access
Independent Use	**Public Interfaces**	**Trusted Intermediary**
Cooperative Use	**Data Pooling**	**Research and Analysis Partnership**
Directed Use	**Prizes & Challenges**	**Intelligence Generation**

Table 1: Data Collaboratives Matrix of Engagement and Accessibility

Current practice in the use of mobility data to address COVID-19 largely aligns with the general data collaboratives typology. Mobility data collaboratives in our sample clustered into six broad categories (in order of prevalence in our sample):

- **Intelligence Generation:** Companies internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public. There are 12 intelligence generation data collaboratives in the sample.
    - Example: Using data provided by the Chinese search engine Baidu, this project measures the effectiveness of non-pharmaceutical interventions in response to COVID-19 across China.[43] It analyzed first-party data capturing movement within and outside of cities, proximity and contact patterns, and the timeliness of disease reporting to construct daily travel networks and simulate likely outbreak scenarios. It then made these models, but not the underlying data, available to inform research and public decision-making.
- **Research and Analysis Partnership:** Companies engage directly with public sector partners and share certain proprietary data assets to generate new knowledge with public value. There are 20 research and analysis partnership data collaboratives in the sample.
    - Example: The COVID-19 Mobility Project is an effort from the Robert Koch Institute and Humboldt University of Berlin to use telecom data from the Deutsche Telekom (distributed by the company Motionlogic) and Telefónica (analyzed and aggregated by the mobility data company Teralytics) to improve understanding of changes in mobility throughout the COVID-19 pandemic.[44] The project maintains a "mobility monitor" which displays daily changes in the number of trips across Germany from March onward. It also produces reports describing major findings following notable changes.
- **Public Interface:** Companies provide open access to certain data assets, enabling independent

42. The Explorer features over 200 data collaboratives as of December 2020. https://datacollaboratives.org/explorer.html
43. 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 17, 2020): 410–13. https://doi.org/10.1038/s41586-020-2293-x.
44. Robert Koch Institute & Humboldt University of Berlin. "Covid-19 Mobility Project." Covid-19 Mobility Project, 2020. https://www.covid-19-mobility.org/.

uses of the data by external parties. There are nine public interface data collaboratives in the sample.

- Example: For its COVID-19 Data Consortium,[45] Safegraph has aggregated foot traffic data in North America, with over 6 million points of interest (POI) across the country. This dashboard visualizes the impact the pandemic has had on the US economy and daily life by comparing data from recent months on commercial activity (foot traffic data) to that of previous years.

- **Trusted Intermediary:** Third-party actors, such as non-profit organizations or humanitarian agencies, support collaboration between private sector data providers and data users from the public sector, civil society, or academia. There are three trusted intermediary data collaboratives in the sample.
    - Example: Flowminder, a data analytics nonprofit, partnered with Ghana Statistical Services to analyze de-identified and aggregated CDRs provided by Vodafone.[46] The effort is aimed at identifying where mobility restrictions instituted by the government are being adhered to on a district, regional, and national level.
- **Data Pooling:** Companies and other data holders agree to create a unified presentation of datasets as a collection accessible by multiple parties. There are six data pooling data collaboratives in the sample.
    - Example: UNICEF, through its big data platform Magic Box, is working to measure the secondary effects of COVID-19 and response efforts, to understand the impacts on social behavior, education, critical supplies, sentiment, opinion, and vulnerable populations. It includes data on the "timely monitoring of physical distancing, evidence on the suitability and sustainability of mobility reductions for low income settings, and better models that allow a better understanding and balancing of the potential impact of these measures on the disease as well as on the underlying communities."[47]
- **Challenge Prize:** Companies make data available to participants who compete to develop apps; answer problem statements; test hypotheses and premises; or pioneer innovative uses of data for the public interest and to provide business value. There are two challenge prize data collaboratives in the sample.
    - Example: The #MOVID19 Hackathon was a competition hosted by Datasketch, a data, software, and mapping business, the New Urban Mobility Alliance, a global technology and mobility organization, and several other local civic organizations.[48] The challenge asked participants to develop innovative analyses of the COVID-19 pandemic and Bogotá's quarantine. Participants used open mobility data from city transit authorities from the sponsor organizations hosted on Github. 44 people contributed, with three projects awarded small cash prizes. These projects included a digital tool to reroute transit and an effort to identify which clinics needed more mobility options.

Challenge prizes are the least prevalent collaborative type in our extensive but not comprehensive sample based on research conducted in English exclusively. We speculate this disparity could be the result of difficulties in balancing the utility of mobility data (and thus its sensitivity) and the protections required to make it accessible to distributed challenge participants as there are many examples of challenge prizes in The GovLab's Data Collaboratives Explorer, which collects projects beyond the context of the pandemic.[49] This disparity could be a signal that the difficulties of making

45. Safegraph. "The Impact of Coronavirus (COVID-19) on Foot Traffic." SafeGraph, December 23, 2020. https://safegraph.com/data-examples/covid19-commerce-patterns/.
46. Ghana COVID 19 Hub. "Ghana COVID 19 Hub Main Page," 2020. https://ghcovid19-statsghana.hub.arcgis.com/.
47. UNICEF Office of Innovation. "Effects of Physical Distancing Measures." UNICEF, 2020. https://www.unicef.org/innovation/magicbox/covid.
48. GitHub. "Datasketch/MOVID19: #MOVID19 Hackaton Para Buscar Soluciones al Reto Del Covid-19 En Transporte Público En Bogotá," 2020. https://github.com/datasketch/MOVID19.
49. Data Collaboratives. "Data Collaboratives Explorer," 2018. http://datacollaboratives.org/explorer.html.

mobility available to challenge participants can be navigated, but that could require a longer lead time than would be workable for rapidly emerging situations such as the COVID-19.

3. Current Practice: Insights from Five Case Studies

The previous sections of this document establish how mobility data has been used by businesses, researchers, and public sector agencies. They show mobility data possesses certain inherent risks that require competent management, institutional safeguards, and personnel committed to protecting data subjects and their rights. However, these sections also show that mobility data, when used responsibly, can provide valuable insights that support public health objectives. Indeed, the projects referenced above suggest governments, research institutions, civil society groups, and businesses consider location data to be useful for COVID-19 response and recovery. While the outcome of many of these projects remain unclear, there are already organizations applying location data to COVID-19 around the globe.

To better understand the elements of this ongoing work around mobility data, the research team researched five entries in its mobility data collaborative repository. While the researchers preferenced projects where they could secure interviews, these case studies explored projects launched in different regions that used different data collaborative types, project scopes, and data types. By drawing on these case studies, together with examples from the broader repository, the team seeks to provide insights on the value, needs, obstacles and preconditions for success of data collaboratives while also highlighting governance and data responsibility challenges involved in using mobility data for COVID-19 response and beyond. While data collaboratives often involve multiple parties, we focus on the experiences of the demand side (data users) and the supply side (data providers).

3.A. Comparative Insights

Table 2: Overview of Insights from Mobility Data Collaborative Case Studies

	Description	Intended and Realized Impacts	Enablers	Risks and Challenges
COVID-19 Mobility Data Network and Facebook Data for Good	The COVID-19 Mobility Data Network, a data pooling collaborative, merges the efforts of Facebook Data for Good and other mobility data providers and over 100 researchers and epidemiologists to understand how, in lieu of a vaccine, communities around the world are relying on non-pharmaceutical interventions to slow the spread of COVID-19.	The COVID-19 Mobility Network Movement Range Map Dashboard aimed to support local and state governments with county-level insights on patterns in human mobility. The Network mobilized researchers to work on operational health issues and continues to enable policy monitoring efforts, including for crises beyond just the COVID-19 pandemic	The research team noted the publishing of the movement range map data on HDX allowed for a significant increase in uptake of the dataset from decision-makers. The team also noted that participants, particularly researchers, were motivated by a sense of public purpose.	The Network faced some operational challenges in releasing data. For example, it produced datasets too large for researchers to use them. In addition, gaps in data and analytical capacities between the non profit and academic community highlighted a need for increased capacity building and data literacy programs as well as external data expertise to have on call for emerging threats.

	Description	Intended and Realized Impacts	Enablers	Risks and Challenges
European Commission Joint Research Centre – Mobility Functional Area Study	A data pooling collaborative spearheaded by the European Commission Joint Research Centre in collaboration with a dozen telecom companies from across the EU. The effort used aggregated and de-identified origin-destination matrices (ODMs)—the number of trips between certain points during a certain reference period—to define geographic units in each country that correspond with real-world mobility patterns rather than traditional administrative boundaries.	The analysis established a method for identifying Mobility Functional Areas (MFAs) that capture real-world mobility patterns. These MFAs support policymakers in deploying mobility restrictions and other interventions in a more targeted way. To date, MFAs have been integrated into a platform for approved decision-makers across the EU. The Commission has formally recommended their use by member states.	The collaborative was created urgently with little lead time. The telecoms were able to leverage existing data sharing infrastructure, products, and programs to rapidly and securely begin sharing insights with the European Commission. The effort was initiated after a high-level call to action by the European Commissioner for the Internal Market. This senior political buy-in (and pressure) likely helped to accelerate the effort.	The data provided by the telecoms were not uniformly structured. One ODM might capture daily movements between two cities while another could be an hourly view of movements between provinces. These disparities necessitated additional processing to render the ODMs interoperable. To guard against re-identification, ODMs only recorded the number of aggregated movements in an area if they reached a certain threshold. This important safeguard also lessened the mobility signal during peak lockdown periods since fewer people left their homes.
Greater London Authority Datastore – COVID-19 Mobility Report	Led by the Greater London Authority's Datastore—a public repository of freely available datasets related to the city of London—the COVID-19 Mobility Report combines six mobility datasets from mostly private sources to provide insights into changing mobility dynamics in London. Analyses include pedestrian activity, retail visitation, and home working.	The data collaborative provides a holistic view into how mobility has been impacted by COVID-19, while making aggregated datasets publicly available for bespoke analysis and sets the stage for continued use of private sector mobility data for economic recovery analyses.	Expertise in data fusion with the ability to combine multiple datasets in order to provide streamlined reports. The Datastore was a pre-existing platform, enabling reach to a wide audience.	Difficulty in sourcing public sector data from within the same public organism.
Informes de Movilidad – Telefónica Chile and Universidad del Desarrollo	Through a multi-sector data collaborative in Chile, researchers at the Universidad de Desarrollo's Data Science Institute create nation-wide mobility	While researchers initially intended to publish academic reports on mobility in Chile, thanks to widespread media coverage and corporate philanthropy, this project grew to a wider program providing vital situational awareness to Chile's Ministry of Science, and the general public.	This project benefited from a longstanding partnership between Telefonica Chile and UDD. UDD's existing collaborative network of data scientists and epidemiologists also helped accelerate methodology development. Finally, the project was further enabled by public exposure through data journalism.	Since the Informes de Movilidad have a direct impact on public policy, and given Chile's historically polarized political climate, there was a strong need for statistical robustness of the models used to create the reports.

	Description	Intended and Realized Impacts	Enablers	Risks and Challenges
University of Toronto Canadian Mobility Analysis	Mobility Proximity in Canada During the COVID-19 Pandemic is the result of a research analysis partnership data collaborative between the Innovation Policy Lab at the University of Toronto, Cuebiq, and the ISI Foundation. The report highlights changes in mobility and proximity throughout the COVID-19 pandemic in Canada. Using anonymized smartphone data with differential privacy from Cuebiq, the publication helps decision-makers and the public to understand the impacts of social distancing measures on Canadian life. The report found that while movement levels have been largely restored, proximity levels have not increased and Canadians continue to avoid interacting with others.	Researchers intended the work would help policymakers, scholars, and the public understand the impact of restrictions and public adherence to them. Following publication, the report received significant press coverage that the authors of the report believe influenced public behavior and decisions. The research team also presented findings to various local officials through a series of public policy seminars.	The research team credited significant support from its partner organizations, Cuebiq and the ISI Foundation, as being integral to the project. The partners were able to provide technical advice and guidance on analysis that informed the work.	Researchers encountered some time delays due to the need to negotiate a responsible data sharing agreement that met the needs of Cuebiq and the Innovation Policy Lab. Lockdown measures meant that researchers could not easily access university equipment to securely access datasets. The team needed to develop workarounds to address these and other logistical challenges. The team had limited financial resources after it did not receive grant funding.

3.B. Insights from Data Users

As previously detailed, there are a number of different mobility data user personae with distinct needs, capabilities and resources. In our detailed case studies alone, users include academic researchers, local city authorities, national governments, and intergovernmental organizations. Within the broader mobility data collaborative repository, users are further represented by data journalists, NGOs, and the general public. Taken together, we provide here common needs, obstacles, enabling conditions and issues around governance and privacy.

Needs

- **Real-time data:** Given the rapidly evolving nature of the COVID-19 pandemic—with constantly shifting geographic hot spots, heterogeneous containment policies and rampant misinformation within the public realm—the need for consistent and real-time data is paramount. This need is especially true for the public sector. While an abundance of academic literature on mobility has been published throughout the pandemic, static reporting from the beginning of 2020 are of limited utility for decision-makers weighing quarantine and economic reopening policies today. In the case of the Greater London Authority's Datastore, for example, the Datastore team sought regularly updated datasets—such as Google's COVID-19 Community Mobility Reports—to provide updated reports on the dynamic state of mobility in London throughout the city's multiple lockdowns.
- **Locally relevant data:** COVID-19 cases have appeared in nearly every country around the world,

yet the distribution of cases across regions, countries, and even within cities, is far from uniform.[50] Highly aggregated statistics at a national or county-equivalent may be sufficient for certain user personae, such as national health authorities or epidemiologists. For other users like local public officials, however, localized data is often required for contextualized decision-making. While Google's COVID-19 Community Mobility Reports met the real-time needs of the London Datastore, for example, the same data source was deemed insufficient for inclusion in Chile's Informes de Movilidad, since it only provides insights at the provincial level. Similarly, when the City of New York launched its Recovery Data Partnership, New York officials sought out mobility datasets from Safegraph, Foursquare, and Cuebiq, which offer city and borough-level insights needed to inform decision-making in the US's largest city.[51] Localized data is particularly important for cities or local governments in developing countries, where there is traditionally poor access to relevant data and a lack of competencies or skills (e.g. developers, data scientists)

- **Capacity, allies, and intermediaries:** Actors that could represent the demand for mobility data, including public health bodies and government agencies, may lack the internal data science capacity necessary to design and implement the analyzed data made available to them. Several projects in our repository and analyzed in our case studies involved public entities forming consortia or engaging intermediaries to extend their capacity for deriving value from mobility data. The continued development of incentives and business models for data intermediaries could help to unlock more and more responsible use of insights derived from mobility data for public problem-solving.

Obstacles

- **Interoperability:** Singular datasets are rarely capable of delivering transformative insights. Many of the data collaboratives in our sample involve the commingling or pooling of mobility data of multiple types and from multiple sources. Even when the gathered datasets are of the same format, structural differences, such as the data's reference period or geographic unit, or disparities in their approach to metadata could introduce analytical challenges for demand side actors.
- **Shifting focus and priorities:** A data collaborative's effectiveness and level of responsibility can depend on having a clear purpose, organizing questions, or problem statement. Data collaboration can require significant lead time to allow parties to come to terms on operational, governance, legal, technical, and financial considerations. In a rapidly emerging environment, such as a rampant global pandemic, the priority questions or analytical focus areas can rapidly shift due to abrupt changes in political, social, economic, or public health circumstances.
- **Siloed data holders:** When seeking access to relevant datasets with the potential to inform decision-making, a logical place for public sector stakeholders to start is within the same public organism in which they operate. However, as demonstrated across several data collaboratives within the detailed case studies and broader repository, it was sometimes easier for public agencies to source mobility data externally from the private sector than from internal agencies under the same umbrella, such as transportation authorities.
- **Lack of resources:** Even absent the COVID-19 pandemic, many organizations operate in a world of constrained resources. A lack of funding or ability to access necessary equipment can force organizations to stop their research or scale back their scope. It can also force organizations to rely on third party consultants and other groups. Many private-sector data providers have stepped in to open access to mobility datasets through data philanthropy programs. Although some mobility providers had existing corporate social responsibility and "Data for Good" programs before the

50. World Health Organization. "WHO Coronavirus Disease (COVID-19) Dashboard." Accessed January 5, 2021. https://covid19.who.int.
51. NYC Mayor's Office of Data Analytics, *supra* note 29.

pandemic began, it remains unclear to what extent continued in-kind data donations will be the norm across the entire ecosystem.

- **Data Limitations:** Mobility data can be the source of significant insights on human movement but it is not wholly representative of real-world activity. Rather, mobility data relies on proxies, such as the movement of cellphones, smartphones and other devices. CDR originating from a single telco, for example, will only contain a user panel of the company's subscribers, which may skew towards certain demographic segments. As previously mentioned, smartphone-derived mobility data can in some cases under represent elderly and low-income populations. It is therefore necessary for mobility data users to ascertain the representativeness of the mobility data in question, and, if necessary, to perform post stratification to create a properly weighted user panel.[52]

Enabling Conditions for Success

Actionable problem definition: The use of mobility data is unlikely to create decision-making value absent a clear, granular, and actionable problem definition or organizing question. Given the many competing needs and priorities involved in the response to COVID-19, practitioners could look to participatory problem definition efforts that enable experts and citizens to identify the use cases with the greatest potential for positive impact.[53]

- **Clear safeguards:** Mobility data can be highly sensitive. Demand-side actors, especially those subject to institutional review boards or other ethics bodies and processes, require clear articulation from data providers regarding how data is being protected and how it can (and cannot) be used. Without guardrails in place, demand-side actors are likely to struggle both in the design of responsible data reuse efforts and in gaining institutional support and clearance for these efforts.

- **Data competency, expertise and right-sizing:** Demand-side actors will struggle to benefit from mobility data collaboration absent not just the relevant data science expertise required to manipulate draw insight from data, but also data competency among other actors and decision-makers who play a role in setting institutional priorities or putting insights into action. Absent sophisticated data science competencies, demand-side actors can benefit from "right sizing" the data that they consume. For example, if NGOs lack in-house data science expertise, they can seek out and use more aggregated, user-friendly datasets in the form of dashboards.

- **Ability to translate insights into action:** Deriving insights from mobility data for COVID-19 response is no doubt an important endeavor, but in order for data collaboratives to be truly effective, it is important that results do not simply sit on a shelf. By widely broadcasting research findings and actively networking with decision-makers who have the ability to put results into practice, data collaboratives can better position themselves to make a real impact. Practitioners should guard against both misuse of data, as well as missed use of data.[54]

- **Collaborative networks:** Perhaps the strongest enabler of success is the ability for demand-side actors to draw on a collaborative network of data-scientists, researchers, and public sector stakeholders. Among our five detailed case studies and roughly 50 data collaboratives featured in the broader repository, social networks abound, with many data scientists teaming up across different institutions to co-author papers and reports. Such networks facilitate the sharing of methods, data resources, and even funding in order to accelerate the pace of research and downstream impact.

52. Bakker, Michiel, Alex Berke, Matt Groh, Alex'Sandy Pentland, and Esteban Moro. "Effect of social distancing measures in the New York City metropolitan area." *Boston: Massachusetts Institute of Technology* (2020). http://curveflattening.media.mit.edu/socialdistance4_14.pdf
53. Verhulst, Stefaan. "Solving a Problem Starts by Asking the Right Question | Apolitical." apolitical, September 2, 2019. https://apolitical.co/en/solution_article/raw-data-wont-solve-our-problems-asking-the-right-questions-will.
54. https://www.unglobalpulse.org/2019/03/unpacking-the-issue-of-missed-use-and-misuse-of-data/

- **Agile Leadership:** Given the rapidly developing nature of the pandemic, many data collaboratives can benefit from leaders who can quickly coordinate and facilitate data sharing. These individuals, data stewards, can allow organizations to work toward their goals of promoting the public instead of letting work become bogged down in bureaucratic or administrative procedures.

Governance and Privacy Issues

- **Scope of purpose and retention:** The severity of the COVID-19 pandemic calls for an extraordinary multi-sector response. However, what is required during a global emergency may not be good practice during normal times, especially when privacy, civil liberties are at stake. When designing data collaboratives, it is crucial to determine the scope of purpose for the use case while also defining clear guidelines around when underlying data should be destroyed, as well as what use cases or applications should be off limits, such as law enforcement or surveillance of individuals and vulnerable populations.
- **Working with trusted intermediaries:** When demand-side actors include public sector authorities, it can be helpful to work with intermediaries that are trusted (or even appointed) by the general public. These actors can serve in multiple ways, such as ethical review boards, or as true intermediaries who execute all of the data processing before passing on highly aggregated insights to the public sector.
- **Compiling previously distributed mobility datastreams:** Just as the compilation of various data streams from various data sources can create interoperability challenges for demand side actors, pooled data can also amplify privacy risks. For example, certain de-identified datasets could be subject to the "mosaic effect." The mosaic effect refers to the analysis of seemingly innocuous, de-identified data together with other available datastreams to re-identify individuals.[55] The compilation of multiple datastreams could also create risks that differential privacy mechanisms, which introduce synthetic noise into a dataset without decreasing its analytical utility, might falter if the same mobility patterns are captured and made accessible by multiple data providers.

3.C. Insights from Data Suppliers

While each mobility data type has benefits and limitations, a common factor between them is that most are derived from private sector companies. This origin carries with it shared needs, obstacles, and enablers for the successful implementation of mobility data collaboratives from the supply side. It also has implications for governance and privacy.

Needs

- **Regulatory clarity:** While it may appear that privacy regulations such as the California Consumer Privacy Act and GDPR would be anathema to private sector data companies, such regulations are effective at setting clear ground rules and distinguishing between fair players and those who act in bad faith. Indeed, private sector mobility data providers have a vested interest in establishing a clear and fair regulatory environment to enable both long term and ethical value creation.
- **Well-structured agreements:** When establishing collaborative initiatives between data providers and users, it is important for parties to begin with a shared understanding of each other's goals, requirements, and responsibilities. While emergencies such as pandemics call for swift action, well-

55. Office of the Assistant Secretary for Planning and Evaluation. "C. The Mosaic Effect." ASPE, November 23, 2015. https://aspe.hhs.gov/report/minimizing-disclosure-risk-hhs-open-data-initiatives/c-mosaic-effect.

structured agreements are critical not only for operational purposes but also to ensure that demand-side actors understand and observe regulations and applicable laws related to handling mobility data.

Obstacles

- **Perception risks:** Even when data is highly aggregated to preserve privacy and provided philanthropically for COVID-19 response, data suppliers run the risk of being perceived as aiding unpopular government policies—such as lockdowns—especially when such policies are related to controversial or unpopular decisions by government users. Some organizations may also be reluctant to initiate a data partnership to mitigate the perceived risk of sharing sensitive data.

- **Long-term sustainability:** With ongoing financial challenges related to the halting economic recovery across the world, many private sector data firms will have to consider the feasibility of continuing to engage in data philanthropy, especially if data philanthropy programs represent a cost center for the business.

Enabling Conditions for Success

- **Existing Corporate Social Responsibility (CSR) program:** Companies with existing CSR programs, such as Telefónica Chile, were well prepared to supply privacy preserving data for COVID-19 response because the data exchange was an extension of existing business practices. CSR programs and social impact initiatives can allow organizations to develop the competencies needed to responsibly share data when a crisis emerges.

- **Collaborative networks:** In the same way that collaborative networks aid supply-side actors seeking co-collaborators and support on topics such as data science and methodology development, so too do data suppliers benefit from strong collaborative networks that can facilitate data collaboration. For example, when COVID-19 emerged as a global health threat, Cuebiq's existing Data for Good program had developed collaborations with dozens of research institutions and NGOs across North America and Europe. By establishing its own data collaborative in the early days of the pandemic, Cuebiq was able to facilitate methodology sharing and collaboration between existing partners, such as the ISI Foundation, and new collaborators, such as the University of Toronto.

- **Detailed demand definition:** Since many big data companies in the private sector spend their resources on designing products and services for commercial applications, it is helpful for supply-side actors to have a clear understanding of the demand for data among public sector users and their needs. If a demand-side actor expresses the need for user-friendly insights, a data supplier might seek to provide highly aggregated and privacy-preserving insights through more simplistic interfaces, such as dashboards.

Governance and Privacy Issues

Given the privacy sensitivity inherent in human mobility data, clear governance, compliance and privacy frameworks are necessary to ethically use mobility data for socially impactful work, such as COVID-19 response. Such data governance and responsibility frameworks are most effective when they identify and mitigate risks across the data lifecycle, which includes planning, collecting, processing, sharing, analyzing, and using. Opportunities, risks, and challenges can emerge at each stage of the life cycle, and risks introduced at earlier stages can compound and accumulate as the project progresses. Organizations have an obligation to be aware of these consequences.

Just as there are successful cases of mobility data being used for good, there are also instances of unethical practices resulting from the absence of rigorous governance and privacy frameworks. Some of these practices are a consequence of the current mobility data ecosystem, an ecosystem in which mobility data is often shared through data brokerages, a process in which datasets are sourced from a variety of suppliers and then resold to interested third party users. Experts around the world have voiced several concerns over this approach, including its implications for surveillance efforts, the loss of individual privacy, and misuse of personal data. For example, NRK Radio, a Norwegian government-owned radio and television public broadcasting company, released a report indicating that over 8,300 mobile devices were being tracked while at hospitals or women's shelters. This location data was easily available to NRK for purchase (for about USD 4,000), allowing them to gain access to information on the travel history of thousands of Norwegian citizens.[56] In addition, the location-data company X-Mode (which itself conducted COVID-19 related analyses[57]), was found to be collecting precise location data from users of a popular Muslim prayer app without obtaining explicit and named consent from users.[58] Beyond the surreptitious provenance of its data, X-Mode also sold individual-level location data to downstream users including private investigator firms and the US military's counterterrorism operations.[59] While this data was pseudo-anonymized, it did not incorporate differential privacy enhancements to preclude re-identification of individual users.[60]

Interviews and desk research found that individuals who participated in all five of the projects understood the need to guarantee competent and responsible use of mobility data. However, given the dearth of clear national or internationally agreed standards of practice, many institutions are left to develop their own frameworks that rely on existing institutional mechanisms to promote ethical use. This lack of consistent and transparent standards alongside the absence of accountability mechanisms represents a notable gap in several of the projects. The research team notes that data users and data suppliers have an obligation to acknowledge and proactively mitigate risk in the areas they work and to promote meaningful governance and privacy standards.

56. Gundersen, Martin. "My Phone Was Spying on Me, so I Tracked Down the Surveillants." NRKbeta, December 3, 2020. https://nrkbeta.no/2020/12/03/my-phone-was-spying-on-me-so-i-tracked-down-the-surveillants/.
57. X-Mode Blog. "X-Mode Offers Premium Data for COVID-19 Research." Accessed January 7, 2021. https://xmode.io/covid-19/.
58. Ibid.
59. Cox, Joseph. "Apple and Google Push Location Tracker X-Mode Out of App Stores." *Vice*, December 9, 2020, sec. Tech. https://www.vice.com/en/article/g5bjpx/apple-google-x-mode-xmode-app-stores.
60. Cox *supra* note 15.

4. Prescriptive Analysis: Recommendations for the Future

In the previous sections of this document, we describe what mobility data is and how it has been used to produce insights on COVID-19. While mobility data has unique characteristics that differentiate it from other data types, it still fits into a broader framework of data use and reuse, one that The GovLab, Cuebiq, and their partners have used to inform their reaction to the current crisis.

In March 2020, the GovLab put out a call to build the "data infrastructure and ecosystem [organizations] need to tackle pandemics and other dynamic societal and environmental threats." Amplified by the World Health Organization and endorsed by over 400 signatories, the call to action focused on seven action areas, detailed below.[61] We expand upon these action areas to provide recommendations specific to mobility data while also seeking to improve current practices and enable future collaboration beyond COVID-19.

The following actions have the potential to enable more effective, sustainable, and responsible re-use of mobility data through data collaboration to support decision-making regarding pandemic prevention, monitoring, and response. These actions, recommended in response to the our findings, include:

1. ***Developing and Clarifying Governance Framework*** to enable the trusted, transparent, and accountable reuse of privately held data in the public interest under a clear regulatory framework.

   By creating a common and publicly available mobility data-sharing agreement template, organizations can set and define standards on critical issues such as scope of purpose, data reuse, and prohibited use cases. In coordination with stakeholders (such as public sector agencies), they can also establish standards around privacy-preserving measures that go beyond simple anonymization or outline requirements needed for the specific context. In the Global South, for instance, organizations may outline procedures that encourage Global South institutions to conduct their own data collection and analysis to avoid the need for other countries to extract data.

2. ***Building Capacity of Organizations in the public*** and private sector to reuse and act on data through investments in training, education, and reskilling of relevant authorities; especially driving support for institutions in the Global South

   Organizations on the supply and demand side of data collaboratives could benefit from a focused effort to build the data science skills and capacity they possess. This work could involve retraining existing personnel to bolster their data skills, creating and hiring for new data science roles and functions, and investing in efforts to distribute data science knowledge and competency across the organization instead of creating siloed data teams or units. Building this competency could an organization shift toward a more data-driven culture, and enable it to more rapidly act upon opportunities to reuse data, including mobility data, to respond to emerging situations. Enabling support for traditionally under-resourced institutions, especially those in the Global South, to

61. UN Stats Hub. "Call for Action: COVID-19 Data Collaboratives | COVID-19 Response," March 30, 2020. https://covid-19-response.unstatshub.org/useful-links/covid-19-data-collaboratives/.

self-sufficiently conduct data collection and analysis and can drive a more interconnected, well-equipped data ecosystem.

3. ***Establishing Data Stewards*** in organizations who can coordinate and collaborate with counterparts on using data in the public's interest and acting on it.

In data collaborative efforts in which both a large number of datasets and participants are involved, projects suffer from a lack of clearly defined roles and responsibilities. Data stewards—individuals or teams tasked with collaborating on data projects with external partners, protecting potential customers and users, and acting to ensure this data is used responsibly—can facilitate more coordinated and efficient data projects for emerging threats. [62]

4. ***Establishing Dedicated and Sustainable CSR Programs on Data*** in organizations to coordinate and collaborate with counterparts on using and acting upon data in the public's interest.

By building CSR into the mission of private sector mobility data organizations, it is possible to lay the groundwork for more rapid, effective and ethical deployment of mobility data in emergency situations. Rather than establishing them as one-off projects or cost centers within organizations, corporate social responsibility programs are designed with long-term sustainability in mind, with opportunities for creating shared value both for society, and for the organization.

5. ***Building a Network*** of data stewards to coordinate and streamline efforts while promoting greater transparency; as well as exchange best practices and lessons learned.

Mature networks of collaborators serve as enablers for both demand- and supply-side actors throughout the pandemic. Although some forums for mobility data collaboration already exist, they tend to be sporadic and limited in sector diversity.[63] More consistent collaborative forums composed of players across private, public, and academic sectors have the potential to more rapidly replicate the organic growth of collaborative social networks.

6. ***Engaging Citizens*** about how their data is being used so clearly articulate how they want their data to be responsibly used, shared, and protected.

Before the pandemic, data reuse sparked significant concerns and misconceptions about how data was (and was not) collected, shared, and reused. The urgency and confusion caused by the COVID-19 pandemic has only amplified these challenges. Stakeholders on the supply and demand side of mobility data collaboratives could alleviate public concerns and better reflect public preferences in their work by implementing programs to engage with citizens and other stakeholders. This work could involve organizing public deliberations on the reuse of mobility data through citizens assemblies or collaborative policy development initiatives (also known as crowdlaw).[64] The GovLab's Data Assembly, for example, engaged cohorts of data holders and policymakers, civic rights and advocacy organization representatives, and New Yorkers from across the five boroughs to understand and synthesize these groups' common and competing perceptions on responsible data reuse and COVID-19.[65]

7. ***Promoting Technological Innovation*** through collaboration between funders (e.g. governments and foundations) and researchers (e.g. data scientists) to develop and deploy useful, privacy-preserving technologies.

62. Verhulst, Stefaan. "Wanted: Data Stewards." Brooklyn, New York: The GovLab, March 2020. https://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf.
63. NetMob. "NetMob 2019," July 10, 2019. https://netmob.org/.
64. The GovLab. "CrowdLaw: Online Public Participation in Lawmaking." CrowdLaw-Online Public Participation in Lawmaking. Accessed January 7, 2021. https://crowd.law/.
65. Young, Andrew, Stefaan Verhulst, Nadiya Safonova, and Andrew J. Zahuranec. "The Data Assembly: Responsible Data Re-Use Framework." Brooklyn, New York: The GovLab, n.d. https://thedataassembly.org/files/nyc-data-assembly-report.pdf.

The approaches for privacy preservation and ethical data collection listed in this report represent practices currently deployed in the market broadly or among early adopters. However, with the emergence of compelling new technologies, there is potential to redefine not only how data is collected and shared but also how value is shared across individuals, organizations, and society at large. Technologies such as smart contracts and homomorphic encryption, for example, allow for data owners to retain full control over how their data is used, and what it is used for, while also enabling data scientists to query data without taking control of the first party data itself. These technologies, along with emerging data sharing frameworks, also enable data owners to share in the monetary value generated by their personal data downstream, thereby providing both greater agency and incentive for users to contribute to and engage with the data economy.

8. ***Unlocking Funds*** from a variety of sources to ensure projects are sustainable and can operate long term.

The COVID-19 pandemic has demonstrated unprecedented demand for big data and the potential for the public sector to rapidly adopt emerging technologies in times of crisis. Where possible, the public sector could allocate resources not only for future preparedness but also to build the human resource capacity required to use big data across a multitude of use cases beyond COVID-19 and public health. While some corporations already engage in data philanthropy, the private sector might consider direct funding of readiness and capacity building programs, which would not only generate positive social impact, but also further increase demand for big data for commercial applications within the public sector.

9. ***Increase Research and Spur Evidence Gathering*** by publishing easily accessible research and creating dedicated centers to develop best practices.

Though there is an increasing interest in data collaboration and data reuse, the field is held back, ironically, by a lack of good evidence on the conditions under which these approaches work best. There are few empirical studies that could guide or accelerate new initiatives. Best practices largely remain limited. These gaps make it hard to design good initiatives and to convince organizations to participate and support them. As such, data scientists and researchers might seek to foster new research and to ensure these resources are open access so they can reach a broad audience. A centralized research center, such as an observatory dedicated to assessing the value of mobility data for good, might also be valuable for researchers.

In this report we discuss various enablers and preconditions for successful and effective data collaboratives. Sadly, the greatest enabler of these data collaboratives' success was perhaps the extreme urgency of a tragic global health emergency. By learning from these experiences, however, it is possible to drive more effective collaboration and deliver more meaningful outcomes through the responsible use of big data not only for the current pandemic and recovery process but for a multitude of challenges where big data can be ethically applied for the benefit of society.

5. About the Authors

GovLab

Aditi Ramesh
Andrew Young
Andrew Zahuranec
Stefaan Verhulst

Cuebiq

Brennan Lake

Open Data Institute

Ben Snaith
Olivier Thereaux

The GovLab

The Governance Lab at New York University's Tandon School of Engineering, The GovLab, is an action research center whose mission is to strengthen the ability of institutions – including but not limited to governments – and people to work more openly, collaboratively, effectively and legitimately to make better decisions and solve public problems. The GovLab's existing Data Collaborative framework identifies the factors that make data-sharing partnerships effective in generating public value, as well as providing a blueprint for designing successful Data Collaboratives while also embedding principles of responsible data use. As an extension of this work, The GovLab's #Data4COVID-19 project provides a living repository of 175 multi-stakeholder data collaboratives focusing on data sharing throughout the pandemic, including 31 projects focused primarily on mobility data sharing - enabling a jumpstart of the above project with early deliverables from mid November onwards.

Cuebiq Data for Good

"Data for Good" is the corporate social responsibility program of Cuebiq, a mobility data platform. Since 2016, Cuebiq has provided mobility data and insights for over 100 academic research projects and humanitarian initiatives related to human mobility. Throughout the COVID-19 pandemic alone, Cuebiq established a COVID-19 Data Collaborative to provide privacy preserving mobility data to dozens of organizations, including the CDC, the City of New York, UNICEF, the World Bank, Johns Hopkins University, and Oxford, among many others. Beyond the establishment of a formal COVID-19 Data Collaborative, Cuebiq has built a model for responsible mobility data sharing grounded in ethical data provenance and privacy by design.

Open Data Institute

The ODI was co-founded in 2012 by the inventor of the web Sir Tim Berners-Lee and artificial intelligence expert Sir Nigel Shadbolt to show the value of open data, and to advocate for the innovative use of open data to affect positive change across the globe.

We're an independent, non-profit, non-partisan company that, since our creation, has welcomed high-profile board members including Mumsnet founder Justine Roberts, Lastminute.com founder Baroness Martha Lane Fox and former European Commissioner Neelie Kroes.

Headquartered in London, with an international reach, hundreds of members, thousands of people trained, dozens of startups incubated, and a convening space in the heart of the capital, we invite everyone interested in developing with data – whether on an individual, organisational or global level – to get in touch.

Acknowledgements

The authors would like to thank the following individuals for their support in supporting the interviews, peer review, and report development processes:

Claire Molinaro (Cuebiq)
Michelle Winowatan (GovLab)
Leo Ferres (Universidad del Desarrollo)
Anita McGahan (University of Toronto)
Gabriel Cavalli (University of Toronto)
Daniel Trabucchi (Politecnico di Milano)
Dominique Diouf (Open AIR)
Tarnjeet Kang (Independent Consultant)
Marjiana Novak (Circle Economy)
Albert Ali Salah (Utrecht University)
Laura Gorman (Facebook)
Andrew Schroeder (DirectRelief)
Richard Benjamins (Telefonica Research)
Michele Vespe (European Commission)

6. Appendix 1: Mobility Data Type

Refers to the type of mobility data and its provenance. Its categories include:

Call Detail Records "CDR" (Telco):

Call Detail Records, "CDRs" are records that are generated by telecommunication companies when mobile phone users make calls, send SMS text messages or use data to connect to the internet (see also XDR). When one user calls another, for example, a typical CDR will include a row of data for the following columns:

- Timestamp
- Duration of call
- Caller identifier
- Recipient identifier
- Cell tower IDs (origin and destination)
- Event type (e.g. call, text, data)

To extract location data insights from CDRs, analysts can infer a mobile phone's positioning based on the coverage radius of the cell tower that services the mobile phone's call, text, or data activities. When multiple cell towers are used by a single mobile phone, it is possible to triangulate the position of the phone in order to improve accuracy. Telecommunications companies can also measure the time for cellular antennas to reach a mobile phone in addition to signal strength in order to estimate distance in a process called trilateration.

CDR use generally benefits from a large amount of data in terms of overall panel size (user count) since CDR datasets often contain all records from all users within a given telecommunication company's network. As a drawback, since the radius of coverage of cell towers is quite large, CDRs have significantly lower accuracy and precision as compared to alternative data sources, such as GPS.

x-DRs:

x-DRs function similarly to CDRs with the key distinction that x-DRs are exclusively based on the data usage of a mobile phone, rather than calling or SMS text messaging. Since data usage involves one party, there is only one phone identifier.

Software Development Kit derived Smartphone GPS data (SDK):

A Software Development Kit, "SDK" is a piece of software that is integrated into a smartphone application, generally on Android and iOS—the two most prominent smartphone operating systems. While SDKs can be used for many types of features and functionalities, within the realm of location data an SDK enables smartphone apps to collect location data by using the device's own hardware, such as GPS, Bluetooth and Wi-Fi. (For the purpose of this report, we refer to SDK derived Smartphone data as GPS data.)

The usage of SDK location data generally benefits from a higher accuracy (~10 meters) and precision as compared to sources such as CDRs. On the other hand, some sources of accessible SDK location data may have smaller panel sizes since they are derived from limited sets of applications with various user counts.

An exception to this is SDK derived data from large app developers, such as Facebook (which manages Facebook App, Instagram, and WhatsApp) and Google (which manages Google Maps and YouTube among other platforms), whose apps have billions of users. These companies however, commonly referred to as “walled gardens”, have more restrictions and less flexibility in providing access to mobility data for use outside of their own activities.

- **First-Party, SDK-derived data:**

  “First-party” refers to data that is collected by an entity directly from the smartphone user, without intermediary brokers. This distinction is relevant in regard to both data accuracy, as well as privacy. Since there are no intermediaries collecting the data between the user and the collector, data is generally collected through a single collection methodology and it therefore does not require cleansing or manipulation to produce a methodologically consistent dataset. In terms of privacy, when there is a direct relationship between a user and a collector, there is a transparent path for users to opt-in and understand the extent to which their data will be used downstream, and the direct relationship facilitates GDPR compliance in order to execute user requests such as opting out, portability and erasure.

- **Third-party, SDK-derived data:**

  “Third-party” refers to data that is collected by one entity then resold to others who often combine datasets from various sources. Using data that is collected from multiple sources can make it difficult to conduct methodologically consistent analyses. Additionally, this process obscures the provenance of the data in terms of whether data was provided with informed consent, while also making it difficult to execute GDPR compliance requests.

BidStream Data:

BidStream location data is derived from the bidding that occurs between buyers and sellers of programmatic advertising, such as pay-per-click advertising on websites or mobile apps. When ad impressions are served to a user through programmatic advertising, the advertiser can “bid” on their ad placement through an automated auction process. Even if an advertiser does not win the ad placement, they are able to retrieve metadata from users exposed to the ad, including IP address, which can be used to infer location.

While BidStream data is generally large in scale due to the sheer volume of programmatic advertising, it is typically inaccurate as it infers location through an IP address, which often cover several miles of coverage and thousands of households per IP address. Furthermore, BidStream data presents ethical challenges as the data is obtained without users explicit consent and without a path for users to opt-out from their data being collected.

Wearable GPS:

Wearable devices are accessories that users can physically wear, such as a wristwatch or bracelet, that contain sensors to record users’ physical locations. While some wearable devices pair with a user’s smartphone in order to leverage its GPS capabilities, other devices have standalone GPS hardware installed in them. Wearable GPS data may be made accessible to downstream users by device makers or retained for sole use by the device makers for analytical purposes.

Aerial imagery data:

Aerial mobility data refers to data collected about physical movements of people based on images derived from aerial photography. These images can be derived from means such as satellites and drones.[66] Aerial imagery data can be used to better understand both large scale population-level trends—such as mapping the location of human settlements where public census data is not available—as well finer grain analyses—such as counting the number of vehicles in a parking lot of a given point of interest. Aerial imagery data collection poses unique privacy challenges in the location data space, since there are no clear opt-in or opt-out paths for the subjects of granular aerial photography.

Vehicle GPS:

Vehicle GPS data is one of many pieces of metadata derived from telematics—the collection of data relating to a particular vehicle, including performance, maintenance status and location. Vehicle GPS data can be derived from portable GPS devices that users install in their vehicles or through on-board GPS hardware built into the vehicle, itself. Vehicle GPS data can be distinguished into two categories: fleet and consumer. While fleet data is derived from commercial vehicles, such as long-haul trucks, consumer data is derived from passenger vehicles. Policies regarding data sharing and accessibility vary by automaker.

Bluetooth:

A signal that allows for short-range wireless communication between electronic devices.[67] Bluetooth can be used to determine when and for how long electronic devices, such as smartphones, come into proximity with other bluetooth-enabled devices. Bluetooth technology may be used to collect location data through the use of "beacons," devices that interact with bluetooth-enabled smartphones to serve messaging, such as advertising, or to collect information, such as number of visitors to a certain point of interest. The use of beacons for data collection generally does not provide users with a clear path to opt in or opt out from data collection, while other smartphone-to-smartphone bluetooth-enabled applications do provide clear paths to opt-in and exercise data rights.

Geotagged Social Media Data

Data connected to videos, photographs, messages, and other content posted on social media containing information on the time and geographic location of the user.[68]

66. https://www.usgs.gov/core-science-systems/national-geospatial-program/imagery
67. https://www.everycrsreport.com/files/2020-05-29_IF11559_f93c95d9673977192f24c0c149cbe28ca7f20e50.pdf
68. https://www.sciencedirect.com/science/article/abs/pii/S0261517716301005?casa_token=tLhh5zlgjhkAAAAA:tW51e5tF-juT-plhzBxkuHsw2-1h79w-lFp_IHBBDh4cN_vaM5CGVd7zx2-uEr8tInGLnz-c

7. Appendix 2: Data4Covid Mobility Repository: Taxonomy of Factors, Design Choices, and Variables

Data Collaborative Type: The operational model or collaboration mechanism enabling the project based on The GovLab's typology of data collaboratives:

- Public Interfaces: Companies provide open access to certain data assets, enabling independent uses of the data by external parties.
- Trusted Intermediary: Third-party actors support collaboration between private sector data providers and data users from the public sector, civil society, or academia.
- Data Pooling: Companies and other data holders agree to create a unified presentation of datasets as a collection accessible by multiple parties.
- Research and Analysis Partnership: Companies engage directly with public sector partners and share certain proprietary data assets to generate new knowledge with public value.
- Prizes & Challenges: Companies make data available to participants who compete to develop apps; answer problem statements; test hypotheses and premises; or pioneer innovative uses of data for the public interest and to provide business value.
- Intelligence Generation: Companies internally develop data-driven analyses, tools, and other resources, and release those insights to the broader public.

Value Proposition: Provides an explanation of the type of question the data project seeks to answer, from one of the following four categories:

- Situational Awareness, how increased access to previously inaccessible information might enable stakeholders across sectors to better understand COVID-19 trends;
- Cause and Effect Analysis, attempts to better understand the drivers and consequences of COVID-19 on public health, society, and the economy. They aim to establish which variables affect the pandemic.
- Prediction, attempts to allow stakeholders to assess future risks, needs, and opportunities.
- Impact Assessment, attempts to determine which, whether, and how institutional interventions, such as new policies or services, affect disease spread or the pandemic's secondary effects.

Sector: A high-level bucketing of the Topic Areas above:

1. Public Health
2. Social
3. Economic

Scope: The range of physical area this data project intends to support:

- Local
- Regional
- National
- Multilateral
- Undefined

Data Supplier: All parties involved in providing data for the data project; this can include one or more of the following:

- Private Sector
- Civil Society
- Research/Academia
- General Public
- Public Sector

Data User: All parties using or benefitting from the data project; this can include one or more of the following:

- Private Sector
- Civil Society
- Research/Academia
- General Public
- Public Sector
- Journalistic Media

Outputs: The output of the data project; this can include one or more of the following:

- Data Repository / Portal
- Data Dashboard / Visualization
- Static Reporting
- Mobile App
- Analytics Platform
- User Tool
- Data journalism

Mobility Data Type: Refers to the type of mobility data and its provenance. Its categories include:

- Call Detail Records “CDR” (Telco)
- XDR (Telco)

- First-Party SDK smartphone-derived GPS
- Third-Party SDK smartphone-derived GPS
- BidStream Data
- Wearable GPS
- Aerial
- Vehicle GPS
- Bluetooth
- Opt-in/Opt-out
- Opt-out only
- No explicit opt in or opt out

Data sharing privacy measures:

- Aggregated
- De-identified
- De-identified with noise
- On-premise federated
- Synthetic Mobility data

Region: The World Bank region the project sought to target. (For example, a project by the United Kingdom's Cambridge University to improve disease detection in Malawi would sort into "Sub-Saharan Africa" and not "Europe and Central Asia"). Regions include:

- East Asia and Pacific
- Europe and Central Asia
- Latin America and the Caribbean
- Middle East and North Africa
- North America
- Sub-Saharan Africa
- South Asia
- Global