COVID-19 Mobility Data Network's Use of Facebook Data for Good Mobility Data

ODI

THE**GOV**LAB

cuebiq

Editor's Note

Cuebiq, involved in the research and drafting of these reports, is a collaborator of the COVID-19 Mobility Data Network. To verify the accuracy of the content, an early draft of this document was made available to Laura McGorman (Facebook) and Andrew Schroeder (DirectRelief). Dominique Diouf (Open AIR) provided a peer review to a later version of this case study.
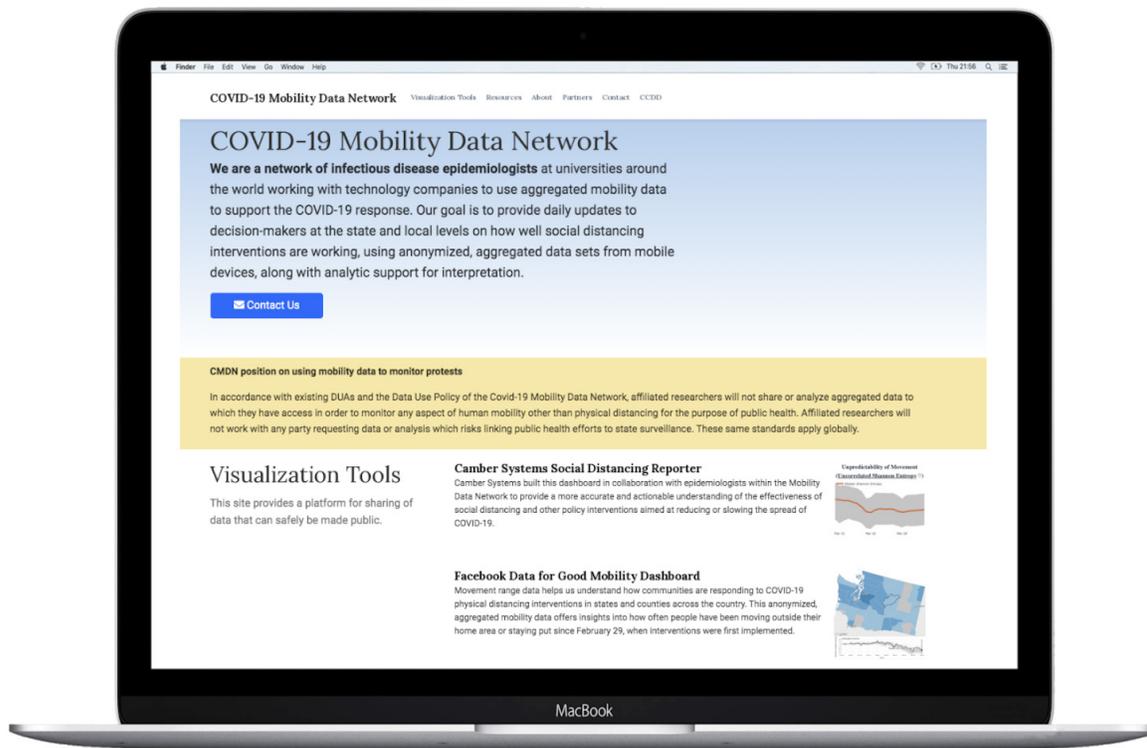
Executive Summary

In April 2020, Facebook expanded its Data for Good offerings to support researchers, international agencies, non-profits, and other public sector institutions in the fight against COVID-19. Facebook's Movement Range Maps, Co-Location Maps, and other Data for Good products incorporate mobility data that, to date, have been used for physical distancing monitoring, identifying high risk areas for disease spread, and other issues. This case study focuses on one of the largest mobility data collaborations created this year, a global data collaboration known as the COVID-19 Mobility Data Network. Data visualizations and reports provided by the Network were used by decision-makers at the national, regional and local level around the world for the purposes of monitoring physical distancing across the country and informing mobility restrictions.

Problem at Hand

Facebook created its Data for Good program in 2017 with a stated mission of supplying non-profit, research, and policy institutions with data for humanitarian response.[1] Facebook launched this program with its first product, Disaster Maps, a self-described effort to make aggregated location signals useful for disaster response efforts. In 2019, Facebook introduced a second product, Disease Prevention Maps, using mobility data to monitor infection disease spread.[2] Since then, partners, including non-profits such as DirectRelief and public research institutions such as the Harvard T.H. Chan School of Public Health, have worked with Facebook for the stated reasons of co-creating and testing data-driven approaches to disaster response, ensuring the accuracy of Data for Good products, and jointly establishing privacy safeguards.[3]

In April 2020, Facebook expanded their Data for Good efforts to provide visualizations, datasets, and surveys to help international agencies, non-profits, and other public sector institutions answer questions related to the COVID-19 pandemic. These new efforts intended to help decision-makers at the regional and local level understand how populations responded to physical distancing orders.[4] At this time, Facebook also partnered with several new organizations to use Disease Prevention Maps to aid relief efforts. The largest of these new collaborations was the COVID-19 Mobility Data Network, which the company describes as "a global coalition to provide real-time insights from Facebook's Data for Good tools."[5]

Project

The COVID-19 Mobility Data Network[6] is "a network of infectious disease epidemiologists at universities around the world working with technology companies to use aggregated mobility data to support the COVID-19 response."[7] This effort was coordinated by Andrew Schroeder at DirectRelief, Caroline Buckee at the Center for Communicable Disease Dynamics at the Harvard T.H. Chan School of Public Health, and Satchit Balsari at Harvard Medical School and Harvard T.H. Chan School of Public Health. The network used pre-existing relationships with policymakers, academic communities, and technology companies to create a mechanism to coordinate disparate data-sharing agreements for pandemic preparedness.

This effort represents a data pool, which The GovLab defines as "a collaboration in which companies and other data holders create a unified presentation of datasets as a collection accessible by multiple parties."[8] Facebook, Cuebiq, and Camber Systems[9] are the core data providers for this project. According to Network collaborators, each partner follows a Data Use Policy, which outlines standards for data use, privacy, and overarching guiding principles related to data governance and ethics.[10] These datasets are then aggregated by affiliated researchers, and shared with policy makers and communities that seek to use them.

---

1.   McGorman, Laura. GovLab <> Facebook - Mobility Case Study Call. Zoom, December 14, 2020.
2.   Ibid.
3.   "Approach." Facebook Data for Good, dataforgood.fb.com/approach/.
4.   McGorman supra note 1.
5.   "Data for Good: New Tools to Help Health Researchers Track and Combat COVID-19." Facebook, 7 Apr. 2020, about.fb.com/news/2020/04/data-for-good/.
6.   The COVID-19 Mobility Data Network later morphed into a larger network, known as CrisisReady, which covers issues beyond Covid-19. See https://crisisready.io/
7.   Covid-19 Mobility Data Network. 13 Apr. 2019, www.covid19mobility.org/.
8.   The GovLab. "Leveraging Private Data for Public Good ." Datacollaboratives.org, datacollaboratives.org/existing-practices.html.
9.   "Mobility Data Network Partners." COVID-19 Mobility Data Network, 14 Apr. 2020, www.covid19mobility.org/partners/.
10. "Data Use Policy." COVID-19 Mobility Data Network, 14 Apr. 2020, www.covid19mobility.org/datausepolicy.

Data Assets

Since 2017, Facebook has relied on mobility data to develop Data for Good products, including Disaster Maps, which uses aggregated location information to determine if communities have evacuated from natural disasters and whether they've been displaced over long periods of time. Launched in 2019, Disease Prevention Maps are described by Facebook as being "designed to help public health organizations close gaps in understanding where people live, how people are moving, and the state of their cellular connectivity, in order to improve the effectiveness of health campaigns and epidemic response."[11] In the early days of the COVID-19 pandemic, Facebook also launched Co-Location Maps, which inform the potential of transmission of disease by human-to-human contact, Movement Range Maps, which measure the effectiveness of physical distancing and stay at home policies at a county level, and the Social Connectedness Index, which measures friendship ties across geographic boundaries. A variety of these products feed into the COVID-19 Mobility Data Network ecosystem.

The backbone for most of Facebook's mobility data come from a subset of users Facebook calls "Location History Users", which are users that have enabled location services on their mobile device and opt-in to allow their location histories to be recorded. To enable products such as "Nearby Friends" or "Find Wi-Fi," users have the option to provide Facebook with this precise location information.[12] Using this data, Facebook then proactively generates datasets that inform an understanding of changes in human mobility occurring in relation to a baseline.[13]

The COVID-19 Mobility Data Network hosts the Facebook Data for Good Mobility Dashboard, a data visualization displaying "county level changes in the rates of mobility, which can allow people to understand the degree to which their communities may be adopting physical distancing efforts."[14] A few months into the start of the collaboration, Facebook applied differential privacy techniques[15] to make Movement Range Maps publicly available on Humanitarian Data Exchange (HDX), an open data sharing platform run by the United Nations Office for the Coordination of Humanitarian Affairs for humanitarians around the world.[16]

Movement range trends are based on aggregated and de-identified data from Location History Users. The dataset shows two key metrics, Change in Movement and Stay Put, which are both calculated using the Bing Tile system. This system draws a map of the world into several thousand 600 by 600 meter square "tiles"; Facebook users' location history data can then be mapped onto each of these days.[17] The Stay Put metric is measured by examining the total number of Facebook users who were only observed in a single Bing Tile for the duration of the day, to understand "how many people were generally staying near or at home."[18] In contrast, Change in Movement seeks to understand "how much less people are moving around since the onset of the coronavirus epidemic"[19] by looking at the total number of tiles that a population appears in over time and comparing it to a pre-pandemic baseline. Both metrics are then aggregated into insights at the regional/county level and incorporate differential privacy to protect against re-identification.

11. "Disease Prevention Maps." Facebook Data for Good, dataforgood.fb.com/tools/disease-prevention-maps/.
12. Herdağdelen, Amaç, et al. "Protecting Privacy in Facebook Mobility Data during the COVID-19 Response." Facebook Research, 3 June 2020, research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response.
13. McGorman supra note 1.
14. "Facebook Data for Good Mobility Dashboard." COVID-19 Mobility Data Network, 14 Apr. 2020, visualization.covid19mobility.org/?-date=2020-12-11&dates=2020-09-11_2020-12-11®ion=WORLD.
15. Differential Privacy is a system in which data is described through patterns or groupings, obscuring information about specific individuals in that dataset.
16. The Centre for Humanitarian Data, centre.humdata.org/what-we-do/.
17. Herdağdelen, Amaç, et al. "Protecting Privacy in Facebook Mobility Data during the COVID-19 Response." Facebook Research, 3 June 2020, research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/
18. Ibid.
19. Ibid.

Design

The COVID-19 Mobility Data Network has a range of participants, including academic institutions such as Boston University and non-profit organizations such as Flowminder. As described on its website, "[t]he Network makes available trained scientists to serve as trusted intermediaries between the aggregated data received from technology companies and the policy makers and communities that seek to use them." These affiliated researchers use this data to then generate analyses, visualizations and reports on behalf of the Network, from daily situation reports to a mobility monitoring dashboard.

While some Data for Good products, such as Movement Range Maps, are open to the public, others are only available upon request and for specific purposes. These non-public datasets are released through data sharing agreements established with non-profits, public agencies, or universities and are meant to address the specific needs of organizations providing support during crisis events. As a result, Facebook states the "research and development that goes into the maps has been (and continues to be) informed by discussions between Facebook researchers and individuals at the humanitarian organizations."[20] Facebook's mobility data is collected only from users who have location history enabled, which is not always representative of real-time mobility patterns, especially in developing country settings. To mitigate this risk, Facebook claims it makes partner organizations aware of these limitations and also makes an effort to use them in combination with other sources of data. [21]

Collaborators of the COVID-19 Mobility Data Network also stated that they made efforts to mitigate harms to data subjects and project participants. Early in the creation of Network, leading partners created a clear statement of purpose to delineate the need and purpose for the collaboration.[22] The Network created a technical guidance document which outlines "aggregation methods and sources of uncertainty, considerations for public health use and privacy, and methods for analysis of aggregated data and their purpose,"[23] as stated on the COVID-19 Mobility Network website. It is unclear to the the authors of this case study how or to what extent this document was enforced.

It took nearly six months to establish guiding principles, governance structures, and a sustainable model for the COVID-19 Mobility Data Network. The Network faced some operational challenges; for example, data providers did not have previous extensive experience releasing large sets of data at once, so it took the Network some time to build these processes and practices internally.[24] However the Network's efforts soon became self-sustaining, requiring minimal upkeep and oversight from the Network itself and its partners. At the end of the summer, researchers in the Network began returning to their previous research and teaching duties.

Outputs and Intended Impact

Intended Impact

During COVID-19, mobility data has been used to study disease transmission as well as adherence to public policies.[25] According to its collaborators, Facebook's mobility data enabled decision-makers within local and regional governments to pursue more targeted movement restrictions throughout the pandemic.[26] Participants further

20. Maas, Paige, et al. "Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery." Facebook Disaster Maps, May 2019, research.fb.com/wp-content/uploads/2019/04/iscram19_camera_ready.pdf.
21. Ibid.
22. Andrew, Schroeder. COVID-19 Mobility Data Network Interview. Zoom, December 9, 2020.
23. Covid-19 Mobility Data Network. 13 Apr. 2019, www.covid19mobility.org/.
24. Schroeder supra note 22.
25. Buckee, Caroline O., et al. "Aggregated Mobility Data Could Help Fight COVID-19." Science, American Association for the Advancement of Science, 10 Apr. 2020, science.sciencemag.org/content/368/6487/145.2.
26. Schroeder supra note 22.

stated that the Movement Range dataset in particular provides insight for questions such as:

- "What is the reaction of a certain population after stay at home orders were imposed?"

- "Once this population was released from these orders, did mobility levels return normal?"

- "What sub-groups of individuals continued to commute during this time?"[27]

Outputs and Realized Impact

According to interviewees who participated in the effort, mobility data shared through the network and other Data for Good products contributed to policy monitoring around the world. In the early days of California's stay-at-home orders, Facebook's movement range data showed that counties with parks and beaches were showing high rates of movement,[28] which, according to Laura McGorman, Facebook's Data for Good Public Policy Lead, subsequently informed state policies for the closure of these sorts of public settings.[29] As recently as December 2020, Taiwan's president Tsai Ing-wen highlighted the importance of sharing mobility data and Facebook's "important role in Taiwan's epidemic prevention efforts, which has contributed to a stable domestic economy."[30] McGorman described decision-makers' use of these mobility data products to drive public policy by saying:

> "Having that validation from not just the partners but the decision-makers is critical. It's a whole value chain from data providers to partners, who are creating reports or research papers to then delivering those to a team in a decision-making body. We're trying to get to the very end of that chain and get signals that our datasets were in fact helpful and that policymakers made different decisions based on it."[31]

While not the only way to study the aforementioned research questions, researchers who used Movement Range Maps claim the resource allowed them to study complex research questions related to mobility, such as understanding urban transportation systems or disease outbreaks to also understanding changing cultures in the context of the pandemic. In the Netherlands, for example, the data set helped one group study the emergence of bike lanes across Europe.[32] These datasets also enabled researchers within the Network to study the use and applications of mobility data broadly. Caroline Buckee, for example, published the paper Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis in September 2020, arguing that "principles of privacy and data protection are vital in assessing more technical aspects of aggregation and should be an important central feature to guide partnerships with governments who make use of research products." [33]

The COVID-19 Mobility Network further allowed research communities to work on operational public health issues, linking the efforts of practitioners and researchers around the world.[34] For example, the Paris School of Economics had researchers who were already interested in the economic impacts of lockdowns on neighborhoods. When they joined the network, they were able to build upon prior research work while simultaneously furthering public policy agendas.[35]

Finally, the reputational value that the COVID-19 Mobility Data Network provides to data holders should not be ignored.

---

27. Ibid.
28. Aral, Sinan. If States Don't Coordinate Lockdown Procedures Immediately, Our Future Is Grim. Los Angeles Times, 9 Aug. 2020, www.latimes.com/opinion/story/2020-08-09/coronavirus-spread-patterns-social-influence.
29. McGorman supra note 1.
30. Chia-hao, Wu, and Matthew Mazzetta. Facebook an 'Important Partner' in Taiwan's Public Health Efforts: Tsai. Focus Taiwan, 3 Dec. 2020, focustaiwan.tw/sci-tech/202012030019.
31. McGorman supra note 1.
32. McGorman supra note 1.
33. Kishore, Nishant, et al. "Measuring Mobility to Monitor Travel and Physical Distancing Interventions: a Common Framework for Mobile Phone Data Analysis." The Lancet, 1 Nov. 2020, www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30193-X/fulltext.
34. Schroeder supra note 22.
35. Ibid.

As large data providers like Facebook face scrutiny over their practices and the ways their data can be abused, philanthropic ventures such as the COVID-19 Mobility Network provide ways for data providers to argue their work can support research and improve lives in a privacy-preserving way. Projects may facilitate the social good, as researchers claim, but they exist within a broader context that should not be ignored.

Lessons Learned

Enabling Conditions

- **The importance of trusted intermediaries:** The Humanitarian Data Exchange platform enabled broader dissemination of Facebook's datasets.[36] According to Andrew Schroeder, using an intermediary such as HDX that is trusted by researchers or the public enables standardized data processing and ethical review practices. This partnership with HDX also allows Facebook to provide data without it being accused (rightly or wrongly) of choosing who it engages with and what research it decides to approve. Costa Rican government officials, for example, already familiar with the HDX platform, decided to incorporate the movement range maps in the creation of their own policies pertaining to the pandemic.[37]

- **Public good as a motivator:** In his interview, Andrew Schroeder described how participants he interacted with through the COVID-19 Mobility Data Network were motivated by a sense of public purpose. People that don't normally participate in operational public health research efforts suddenly received a model to do so, he argued. Schroeder claimed researchers and participants of the network were incentivized to contribute their time and efforts because they believed they were contributing to an effort that would improve public well-being.[38]

Challenges

The need for a sustainable business model: At the end of summer 2020, according to the collaborators, it became clear to members of the Network that a sustainable business model is required for these efforts to endure. For example, Andrew Schroeder argued for a model in which infectious disease epidemiologists are compensated for their time to work on the Network alongside their own research projects.[39] In addition, while data providers made exceptions for COVID-19 to contribute time and data assets for pandemic relief, this practice is not sustainable without a viable long-term business model.

Compiling previously distributed mobility datastreams: Generally, mobility data is generated as a byproduct of a user releasing consent to a mobile application or social media platform. As Andrew Schroeder noted in interviews, mobility data is unique because it allows researchers to "monitor change in population near real time at scale."[40] This also, however, creates a challenge in sharing mobility data as various providers stack up different sources of data in a collaborative project such as the Network, it often erodes corrections made for differential privacy noise in a single dataset. Data scientists have still yet to uncover the implications in the layering of this data, and these processes should be examined further in the future.

Lack of visibility into the full lifecycle of data: Laura McGorman described the need for greater visibility into all

---

36. Ibid.
37. Ibid.
38. bid.
39. Ibid.
40. Ibid.

aspects of data products. Many participants in the Network had a limited understanding of decision-points along the data lifecycle, and after having delivered reports or analyses to some officials, it was ultimately unclear to even members of the research community how certain analyses affected decision-making down the line.

The need to drive capacity building and data literacy efforts for emerging threats: Some data providers, prior to joining the Network, lacked experience releasing large sets of data at once to the public. As a result, the Network faced some operational challenges in releasing data products at its outset. In addition, according to the Network collaborators, there were gaps in data and analytical capacities between nonprofit and academic communities, highlighting a need for both increased capacity building and data literacy programs, as well as data expertise escrow to have on call of emerging threats. Laura McGorman noted that more work can be done to communicate differences between mobility data products to external partners to enable easier uptake of these datasets.

The need to standardize data responsibility practices: Andrew Schroeder noted the lack of standardized data responsibility practices across the industry, and the increasing need to do so. In his interview, he said, "everyone has to come up with some kind of viable Data for Good model. There should be some kind of industry-wide set of standards and agreements and practices in place that say 'this is how you do this' [...] But, you know, I don't think that really exists yet." Clear guardrails can help establish standardized processes, enhance accountability, and allow data collaboration efforts to gain greater public and institutional support.

Conclusion / Next Steps

According to interviewees who participated in the effort, the technical, legal, and operational infrastructure created by the Network set a precedent for disaster response data sharing for other crises in 2020. The practice of streamlined data collaboration cemented by the creation of the Network allowed participants to also jointly focus on other efforts, such as supporting the state government's response to the California wildfires and hurricanes on the Gulf Coast. Going forward, interviewees noted, sustained involvement from private sector actors, researchers, and non-profit organizations can support more effective data-driven response to emerging crises, from both an intellectual and operational perspective.