

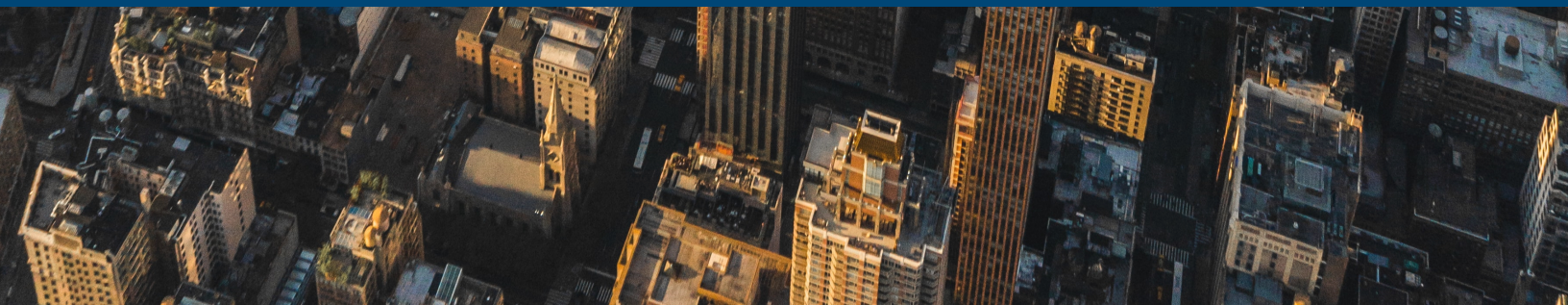


THEGOVLAB



Open Data Policy Lab

NYC RECOVERY DATA PARTNERSHIP

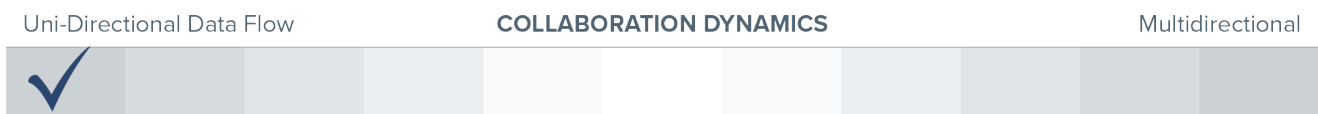
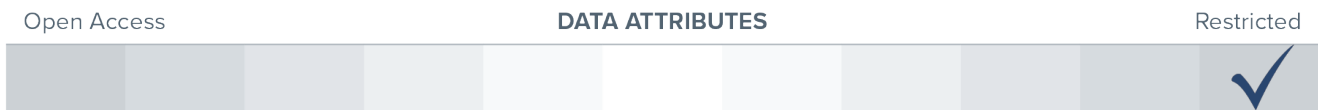


EXECUTIVE SUMMARY

In July 2020, following severe economic and social losses due to the COVID-19 pandemic, the administration of New York City Mayor Bill de Blasio announced the NYC Recovery Data Partnership. This data collaborative asked private and civic organizations with assets relevant to New York City to provide their data to the city. Senior city leaders from the First Deputy Mayor's Office, the Mayor's Office of Operations, Mayor's Office of Information Privacy and Mayor's Office of Data Analytics formed an internal coalition which served as trusted intermediaries, assessing agency requests from city agencies to use the data provided and allocating access accordingly. The data informed internal research conducted by various city agencies, including New York City Emergency Management's Recovery Team and the NYC Department of City Planning. The experience reveals the ability of crises to spur innovation, the value of responsiveness from both data users and data suppliers, and the importance of technical capacity, and the value of a network of peers. In terms of challenges, the experience also exposes the limitations of data, the challenges of compiling complex datasets, and the role of resource constraints.

- **Data Collaborative Model:** In The GovLab's typology of data collaboratives, this arrangement is classified as a trusted intermediary, a type of collaborative in which third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia. The NYC Recovery Data Partnership (itself staffed by the Mayor's Office of Data Analytics, First Deputy Mayor's Office, Mayor's Office of Operations and Mayor's Office of Information Privacy) served as an intermediary to coordinate data exchanges between private sector partners and various NYC agencies.
- **Data Stewardship Approach:** New York City's work demonstrates the value of a data steward for coordinating many disparate data activities, including outreach to potential partners and securing buy-in for internal staff.

Cover Photo by [Triston Dunn](#) on [Unsplash](#)



SETTING THE SCENE

New York City was among the first communities in the United States hit by the COVID-19 pandemic and one of the US cities where the crisis made itself most visible. Through March 2020, the number of cases surged from 16 to over 23,000.¹ As local hospitals scrambled to find beds for patients, local and state leaders began instituting restrictions on residents. On March 22nd, Andrew Cuomo, then-Governor of New York, ordered the closure of all non-essential, in-office business operations across the state.²

This lockdown and the continued spread of the disease had a significant impact on public life across the city. By June, an estimated 526,000 households had filed claims for unemployment insurance; roughly 735,000 households lost income as a result of the pandemic.³ According to the Chamber of Commerce, over a quarter of all Brooklyn businesses reported missing their commercial rent payments.⁴

New York City's government began searching for ways to respond to this unprecedented crisis. Officials sought to use their limited resources in a targeted and impactful way without exacerbating inequalities throughout the city.

ESTABLISHING THE COLLABORATIVE

Joe Berkman-Breen, who served as an advisor in the Mayor's Office of Policy and Planning under the First Deputy Mayor, noticed these efforts by city leaders in the early days of the pandemic. As the weeks went on, he noticed that individual city agencies were reaching out to partners in the private sector, universities, and civil society to collect data to supplement their own data assets. They were looking for data that could be used to better understand how COVID-19 was impacting the city in real time and how the city itself could respond.⁵

Berkman-Breen subsequently began working with the Mayor's Office of Data Analytics as well as the Mayor's Office of Information Privacy to see if there was a way to better coordinate

¹ "N.Y.C. Death Toll Hits 365 as Case Count Tops 23,000." *The New York Times*, March 26, 2020, sec. New York. <https://www.nytimes.com/2020/03/26/nyregion/coronavirus-new-york-update.html>.

² Office of New York Governor Andrew M. Cuomo. "Governor Cuomo Issues Guidance on Essential Services Under The 'New York State on PAUSE' Executive Order." NY.gov, March 20, 2020. <https://www.governor.ny.gov/news/governor-cuomo-issues-guidance-essential-services-under-new-york-state-pause-executive-order>.

³ "Policy Minute: Housing Stability and COVID-19 Recovery." Accessed January 22, 2021. <https://furmancenter.org/thestoop/entry/policy-minute-housing-stability-and-covid-19-recovery>.

⁴ Partnership for New York City. "New York City COVID-19 Economic Impact Update," June 2020. <https://pfnyc.org/research/new-york-city-covid-19-economic-impact-update/>.

⁵ Bjornsson, Oliver. 2021. "RE: Interview Re: NYC Recovery Data Partnership Mobility Data Case Study," February 11, 2021.

these one-off partnerships. He engaged with Kelly Jin, then New York City's chief analytics officer, and Adrienne Schmoeker, then Deputy Director of Analytics for New York City, who helped further develop the overall concept in a series of calls with external data holders.⁶ After several of these calls, Jin and Schmoeker outlined a proposal for a programmatic hub that would allow the City to set common standards for the data it acquired, better share access across city agencies, and maintain a unified governance structure for regulating data access.⁷ Berkman-Breen secured support from City Hall via the First Deputy Mayor's Office and developed a standardized legal agreement with data holders.

On July 23 2020, the Office of New York Mayor Bill de Blasio formally announced the result of these conversations: the NYC Recovery Data Partnership.⁸ As the project website characterized it, the partnership called private organizations, non-profits, and community groups to "share data with the City to aid in Covid-19 response and recovery efforts."⁹ The exchange of data, including mobility data, was coordinated by senior leaders in the New York City government at the Mayor's Office of Policy and Planning, Mayor's Office of Data Analytics, Mayor's Office of Operations and the Mayor's Office of Information Privacy. Partners would receive no compensation for their participation but were recognized publicly for their support in these efforts.¹⁰ Oliver Bjornsson, a policy analyst at the Mayor's Office of Data Analytics, would conduct project management with oversight and support from Schmoeker.¹¹

In The GovLab's typology of data collaboratives, this type arrangement is classified as a trusted intermediary, a type of collaborative in which third-party actors support collaboration between private-sector data providers and data users from the public sector, civil society, or academia.¹² The NYC Recovery Data Partnership served as a specialized intermediary, connecting partners to specific government agencies.

⁶ Schmoeker, Adrienne. 2023. Data4COVID Case Study: NYC Recovery Data Partnership. GoogleDoc.

⁷ Ibid; Leger, Matthew. "NYC's Recovery Data Partnership: Cross-Sector, Values-Driven Data Sharing for Pandemic Response." Data-Smart City Solutions, November 17, 2020.

⁸ Office of the Mayor. "Mayor de Blasio Announces Launch of The NYC Recovery Data Partnership With Open Call for Data." The Official Website of the City of New York, July 23, 2020. <http://www1.nyc.gov/office-of-the-mayor/news/542-20/mayor-de-blasio-launch-the-nyc-recovery-data-partnership-open-call-data>.

⁹ NYC Recovery Data Partnership. "Recovery Data Partnership." NYC Analytics, January 11, 2021. <https://web.archive.org/web/20200810143652/https://www1.nyc.gov/site/analytics/initiatives/recovery-data-partnership.page>.

¹⁰ Bjornsson *supra* note 6.

¹¹ Schmoeker *supra* note 7.

¹² 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>.

Data Assets:

Parallel to the formation of the NYC Recovery Data Partnership conceptually, city officials needed to form agreements with data suppliers who could share usable and relevant data to city agencies. Some of those organizations engaged had already approached the City offering to help while others were already in discussions with specific agencies. Still others received formal requests from city officials based around, as characterized by one city leader, “articulated data needs of agencies, along with a desire to represent a variety of major sectors.”¹³

As relayed by one analyst in the Mayor’s Office of Data Analytics, the City conducted due diligence with all of these groups to ensure information was responsibly collected and handled. They noted:

“As part of our legal agreements, partners have to represent that all data provided to the City was collected from individuals with prior consent and, additionally, where possible, that they will make reasonable attempts to notify individuals that their information was being shared with the City. [...] Our agreements require partners to make best efforts not to share PII with the City, and we have a process for protecting any PII included in the data. The Mayor’s Office of Information Privacy, led by the City’s Chief Privacy Officer, drafted our template agreement and reviews each of the data license agreements the City enters into with partners. The data the City receives is transferred using a secure file transfer protocol consistent with standards set by the NYC Department of Information Technology and Telecommunications and is securely stored in accordance with the City’s information security requirements set by the NYC Cyber Command.”

In total, this work yielded 15 partners (listed in Table 1) capable of providing non-traditional datasets that the City felt could yield insight about the spread and impact of the disease. The data collected by all these companies was aggregated and, per city documentation, lacked identifying information on individual users.¹⁴ All data could be accessed through a central data repository accessible to users at relevant city agencies following request and approval.

We classify this data into four types: health data; non-traditional mobility and geolocation data; non-traditional economic data; and non-traditional sentiment data. As the below indicates, most of the Partnership emphasized economic and mobility data indicators.

¹³ Bjornsson *supra* note 6.

¹⁴ Ibid.

Non-Traditional Health Data

The partnership included one dataset capable of describing the physical health of individuals and populations as it related to COVID-19 diagnoses: Kinsa. A health technology company based in California, Kinsa operates smart thermometers and other internet-connected devices. For the initiative, Kinsa would aggregate and transmit data on smart thermometer readings to identify when and where abnormal temperatures spiked.¹⁵ This would, theoretically, provide the city with an early warning of a major outbreak, though it is unclear what the level of adoption of Kinsa products was at the time of the partnership across the city. In 2021, the *New York Times*, reported a partnership between Kinsa and the New York City Department of Health to distribute 100,000 internet-connected thermometers to New York public schools as part of the former's FLUency program.¹⁶

Kinsa collected data directly from smart thermometer users and an app associated with the tool explicitly asks users to "opt in" to sharing precise location data with the company and possible third parties.¹⁷ Documentation provided by the Partnership indicates Kinsa aggregated all data prior to sharing with city agencies. As it does on its public-facing "Kinsa HealthWeather" platform, the company sought to differentiate COVID-19 from other illnesses by determining if the "number of fevers was statistically higher" than expected for influenza-like diseases.¹⁸

Non-Traditional Mobility and Geolocation Data

The partnership also includes three organizations that collected data on the location of a mobile device relative to people and landmarks in the physical world and other devices. These companies included Safegraph, a geospatial data company; Cuebiq, a location-intelligence company; and Foursquare, a search engine.¹⁹

The first, Safegraph, provided data on points of interest and patterns of foot traffic to companies to assist them in improving their visibility and identifying customers.²⁰ For the recovery partnership, Safegraph provided data on notable points of interest where people spent time or money, weekly and monthly patterns of movement, and social distancing at a

¹⁵ NYC City Hall Press Office. 2020. "Mayor de Blasio Announces Launch of The NYC Recovery Data Partnership With Open Call for Data." The Official Website of the City of New York. July 23, 2020. <http://www1.nyc.gov/office-of-the-mayor/news/542-20/mayor-de-blasio-launch-the-nyc-recovery-data-partnership-open-call-data>.

¹⁶ Anthes, Emily. 2021. "New York Turns to Smart Thermometers for Disease Detection in Schools." *The New York Times*, June 6, 2021, sec. Health. <https://www.nytimes.com/2021/06/06/health/coronavirus-kinsa-thermometer-new-york.html>.

¹⁷ Kinsa Health. 2022. "Privacy Principle." Kinsa. 2022. <https://home.kinsahealth.com/privacy>.

¹⁸ Nadia. 2021. "COVID-19 and Kinsa." Kinsa Help Center. February 2, 2021. <https://support.kinsahealth.com/hc/en-us/articles/360056375411-COVID-19-and-Kinsa>; Kinsa Health. 2023. "New York Illness Risk Map." HealthWeather. 2023. <https://healthweather.us/map/new-york-NY>.

¹⁹ NYC Recovery Data Partnership *supra* note 10.

²⁰ SafeGraph. "Home Page." Accessed January 25, 2021. <https://safegraph.com>.

census block group level.²¹ This data received daily updates that was, in turn, shared with city agencies.

According to Safegraph's privacy policy, it obtained this information "from trusted third-party data partners such as mobile application developers, through APIs and other delivery methods."²² Safegraph does not oversee the collection itself. Instead, individual apps decide on their collection practices, practices which are outlined in the privacy policy and legal terms of the apps themselves.

Cuebiq, meanwhile, analyzes offline consumer behaviors for companies and marketers.²³ For the recovery partnership, Cuebiq provided four types of data. First, it offered a measure of the median distance traveled by all devices on a census block level. Second, it offers a count of aggregate visits to certain points of interest across the city. Third, it assessed the percentage of users staying at home "by measuring how many users moved less than 330 feet from home."²⁴ Finally, Cuebiq also provided the Partnership with its "Traveler Analysis" dataset which, according to its website, "allows decision-makers to estimate not only the number of people traveling to a given location from out of state, but also to understand where exactly travelers are coming from, and what percentage are self-quarantining vs. failing to shelter in place."²⁵

Cuebiq collects this data directly from smartphone users who have agreed to share their locations ("opting in") with certain apps. The company's proprietary software development kit lets Cuebiq "know exactly where the data is coming from and can ensure users are providing consent to data collection."²⁶ It does not rely on the collection practices of third party app developers.

The final mobility data contributor is Foursquare, a technology company whose platforms allow users to search and discover new locations²⁷ and share their locations with friends.²⁸ For the recovery partnership, Foursquare shared aggregate foot traffic data broken down by

²¹ "Recovery Data Partnership Data Catalog." City of New York, September 14, 2020.

²² SafeGraph. "Privacy Policy," January 12, 2021. <https://www.safegraph.com/privacy-policy>.

²³ Chan, Lawrence. "Cuebiq on Location Data Advancements and How Privacy Plays a Role." Cuebiq, March 22, 2019. <https://www.cuebiq.com/resource-center/resources/cuebiq-on-location-data-advancements-and-how-privacy-plays-a-role/>.

²⁴ "Recovery Data Partnership Data Catalog." City of New York, September 14, 2020 ; Cuebiq. "Cuebiq's COVID-19 Mobility Index (CMI)." Accessed January 25, 2021. <https://help.cuebiq.com/hc/en-us/articles/360041285051-Cuebiq-s-COVID-19-Mobility-Index-CMI->.

²⁵ Cuebiq. "Traveler Analysis." Accessed January 25, 2021. <https://www.cuebiq.com/visitation-insights-traveler-analysis/>.

²⁶ Chan *supra* note 24.

²⁷ Foursquare. "Trending This Week: New York City," January 18, 2021. <https://foursquare.com/city-guide>.

²⁸ Swarm App. "Home Page," 2021. <https://www.swarmapp.com/>.

borough, category of destination (e.g. arts and entertainment and colleges and museums),²⁹ number of visits detected at a location type, average time spent at a type of location type, and the extent to which users visiting a location type are over 65.³⁰ Foursquare's data updated daily and spanned from February 2020 to present.

Non-Traditional Economic Data

Of the 15 organizations included in the Partnership, eight provided data that could be directly used to describe the economic activity of individuals, groups, and organizations. Indeed, the companies who formed relationships with the city each had distinct ways in which they measured the “pulse” of economic activity. To assess the economic health of neighborhoods, the Partnership had Streeteasy data, which provided “weekly real estate rental and sales metrics aggregated at the neighborhood level” and ioby, which provided “descriptions of community-based crowdfunding projects and donation transactions.” For business health, Mastercard provided aggregated retail spending insights while BetaNYC offered updated lists of businesses, their hours, and other details. For personal finances, Upsolve gave demographic information about New Yorkers signing up for personal bankruptcy services.³¹

The number of organizations involved makes it difficult to detail the practices of all of these organizations in depth. As with the other data types listed, however, all data transmitted to city agencies was aggregated prior to use. Data often came from information uploaded by users directly to their platforms (as was the case with StreetEasy and LinkedIn) or from consumers directly interacting with services provided by different institutions (as was the case with Upsolve). In the case of Mastercard, the data emerged from points of sale from both purchasers and vendors.³²

Non-Traditional Sentiment Data

Finally, the Partnership included several sources that might be used to understand how individuals and groups perceive developments related to COVID-19 and the world around them. These sources to assess social health included SONYC, a New York University initiative to collect audio from around New York City.³³ SONYC's dataset is based around a tagging system where human operators, gathered from the Zooniverse citizen science platform, listen to over 150 million audio recordings gathered from acoustic sensors throughout the city and

²⁹ A full list of category IDs used by Foursquare can be found here: <https://developer.foursquare.com/docs/build-with-foursquare/categories/>

³⁰ “Recovery Data Partnership Data Catalog.” City of New York, September 14, 2020.

³¹ https://data.cityofnewyork.us/w/ut65-is4f/25te-f2tw?cur=ilHrL3qRto_&from=root

³² Mastercard Data & Services. 2021. “Mastercard Geographic Insights Platform Terms and Conditions for Merchants.” Mastercard Data & Services. March 26, 2021. <https://www.mastercardservices.com/en/solutions/mastercard-geographic-insights-platform-terms-and-conditions-merchants>.

³³ “About.” 2015. *Sounds of New York City* (blog). October 2015. <https://wp.nyu.edu/sonyc/>.

tag them based on the kind of noise they hear (e.g. engines, machines, powered saws, alert signals, music, human voices, dogs). The dataset includes no personally identifiable or sensitive information.³⁴

Non-traditional sentiment data sources also included the Metropolitan Museum of Art, which offered aggregated museum attendance information and information on membership activities.³⁵ The researchers encountered no further detail on what this dataset entailed or how it was structured.

Design:

As relayed by Olivier Bjornsson, an analyst at the NYC Mayor’s Office of Data Analytics, the Recovery Data Partnership was designed in coordination with “experts and advocates in the fields of data, technology, and privacy.”³⁶ Data requests followed the following structure.

The NYC Recovery Data Partnership sent out a notice to each city agency informing them of the partnership’s existence and the datasets available through it. Staff at these agencies could then fill out a centralized form, produced through AirTable, to request specific data assets. Analysts would need to indicate the specific research question their work will be targeted toward, which partners’ data they thought would be most useful for this research, how they planned to analyze this data, the duration of the project, and how the findings from this analysis would inform policy or decisions on COVID-19.³⁷ This application received review from each agency’s counsel’s office.³⁸ Once the counsel was signed off, the partnership’s program team assessed the proposal and determined whether to approve or reject it.

These proposals would then be reviewed by those officials in agencies responsible for overseeing the partnership—the Mayor’s Office of Data Analytics, First Deputy Mayor’s Office, and Mayor’s Office of Information Privacy—informed by the guidance provided by the city’s Chief Privacy Officer and the Mayor’s Office of Information Privacy. Assessments of requests

³⁴ Cartwright, Mark, Jason Cramer, Ana Elisa Mendez Mendez, Yu Wang, Ho-Hsiang Wu, Vincent Lostanlen, Magdalena Fuentes, Graham Dove, Charlie Mydlarz, and Justin Salamon. 2020. “SONYC-UST-V2: An Urban Sound Tagging Dataset with Spatiotemporal Context.” *ArXiv Preprint ArXiv:2009.05188*. <https://arxiv.org/pdf/2009.05188v1.pdf>.

³⁵ NYC Open Data. 2022. “Recovery Data Partnership (RDP) Data Partners.” September 23, 2022. https://data.cityofnewyork.us/w/ut65-is4f/25te-f2tw?cur=ilHrL3qRto_&from=root.

³⁶ Bjornsson *supra* note 6.

³⁷ Recovery Data Partnership. “RDP: Data Request Application for Agency Analysts.” Airtable. Accessed January 26, 2021. <https://airtable.com/shr9C9PxUkFmCRJRZ>.

³⁸ Goldberg, Dara, and Melissa Umberger. Interview re: NYC Recovery Data Partnership Mobility Data Case Study. Zoom, January 15, 2021.

were evaluated under the following criteria (with the support of an external advisory board composed of experts in academia and civil society),³⁹ which was shared with applicants:

- Whether the proposed data use is directly tied to the agency's Covid-19 response and recovery efforts;
- Whether the proposed data use appears reasonable and responsible;
- Whether the proposed data use would provide benefit to New Yorkers; and
- Whether the proposed use is consistent with the terms of the license agreement and in accordance with applicable law.⁴⁰

If approved, the authorized analyst would receive access to a centralized repository, containing all the collected data but with access limited to only those specific datasets they had been approved to use for the time allotted by their request.⁴¹ The city facilitated this data sharing through a SharePoint site that provides either direct API access (enabling individuals to read CSVs directly from the site) or allows individuals to download files. Datasets on the SharePoint site were refreshed on an ongoing basis by the RDP team and the partner organization. It was facilitated by NYC's Department of City Planning's Enterprise Data Management Team, who stood up the platform and code that pulled and refreshed data from partners.⁴² All this work ensured a consistent "data pipeline that houses all of the different vendors' information on a cloud service" accessible to those who need it.⁴³

Requesting city agencies also needed to agree to abide by the terms of the city's data license agreement. While agency counsels must provide signed approval during the application process to ensure they are aware of their obligations, the exchanges themselves were only made possible through a citywide agreement. This overall agreement contained requirements and stipulations about appropriate use, compliance measures, and authorship credit.⁴⁴ As indicated above, non-city organizations seeking to contribute data to the partnership, meanwhile, needed to sign a standardized data sharing agreement with the city. This agreement required data suppliers to "collect their data with lawful consent, publicize their participation in the partnership, and provide comprehensive data documentation to the RDP team."⁴⁵ This template agreement is publicly available.⁴⁶

³⁹ The advisory board served as an attempt at intentional transparency design. Schmoeker *supra* note 7.

⁴⁰ Bjornsson *supra* note 6.

⁴¹ Leger *supra* note 6.

⁴² NYC Regional Planning, 1 Feb 2021: https://docs.google.com/document/d/1ikYOLobF-WKxwUn9OB684DeAYNbg-KJ1ST6QAZ_fCMk/edit

⁴³ Goldberg and Umberger *supra* note 27.

⁴⁴ Ibid.

⁴⁵ Leger *supra* note 8.

⁴⁶ City of New York Mayor's Office of Operations. "NYC Recovery Data Partnership Data License Agreement." City of New York, 2020. <https://drive.google.com/file/d/1HpeV-05ENojDn75154QSXrFBpTaCB7nT/view>.

Analysis & Insights

The intent of this effort was to provide city agencies with access to additional data that can improve their decision-making on issues related to COVID-19 response. While all users were city workers, how they used the data and for what purpose could vary significantly according to agency needs. For the purposes of this case study, researchers interviewed representatives of the New York City Emergency Management's Recovery Team and the NYC Department of City Planning. Both agencies had extensive experience collecting, managing, analyzing, and acting upon data insights and sought out an opportunity to expand their assets by applying to use mobility data from Cuebiq, Safegraph, and other participants in the NYC Recovery Data Partnership. While these agencies do not represent the totality of city mobility data usage (nor the most regular users of it), their work is instructive, demonstrating how quickly agencies could use the Partnership to mobilize in response to a crisis.

The New York City Emergency Management's Recovery Team sought to use mobility data and other assets to visualize recovery trends through a dashboard available to city employees and leaders.⁴⁷ This work began over Summer 2020 because, in the words of one city employee, analysts "were curious about how the city was recovering along economic, social, community, transportation, and housing" lines amid declining cases.⁴⁸ The dashboard provided situational awareness of daily movements around New York.

A team within the NYC Department of City Planning, meanwhile, used mobility data from Cuebiq, Foursquare, and Safegraph to inform mid- to long-range planning around various city policy issues such as housing and transportation. This work included using datasets to observe how people interact with central business districts, how people moved around, and the equity implications of who was going to work and who could not.⁴⁹ It also included efforts to understand how the crisis has displaced residents and changed neighborhoods. Describing the crisis as "closer to a Katrina or a 9/11 than a typical migration year," the department intended to use mobility data to determine, with confidence, how the pandemic has changed the composition of neighborhoods so the city can respond to these shifts and adjust services accordingly. To understand these shifts, analysts created "a tool that looks at how people [move] and estimates the number of trips using SafeGraph [...] comparing that year on year to understand what the patterns are."⁵⁰

⁴⁷ Ibid.

⁴⁸ Goldberg and Umberger *supra* note 27.

⁴⁹ Ibid.

⁵⁰ Ibid.

Products from these analyses were not formally released to the public. The Recovery Framework developed by NYC Emergency Management indirectly informed the City's All-Hazards Recovery Framework, though additional details were not provided.⁵¹

OUTCOME AND IMPACT

In the first year after the onset of the pandemic, analysts and data vendors reported the NYC Recovery Data Partnership served as an important source of insight on how the pandemic has affected the city and its residents and businesses.⁵² These insights were not just relevant in the short-term for immediate response work but in the long term as well, informing long-term strategic objectives. As the effects of the crisis linger, it is critical for leaders to understand the challenges New Yorkers face. While neither of the two analysts interviewed for this case study could point to a specific decision affected by the Recovery Data Partnership, they argued it provided critical insights that informed existing policy. Those involved in setting up the system argued that its creation, in and of itself, was a sign of success and demonstrated that large-scale data collaborations were feasible at the time and in the future.⁵³

Indeed, one of the most surprising aspects of the project for its participants was the speed of the initiative. After requesting data, agency analysts received timely responses to their requests. The multi-layer reviews of use case applications tended to take about one week to conduct and complete.⁵⁴ This speed from the partnership's leadership to review proposals, and share data meant agencies could begin work almost immediately instead of being bogged down by usual bureaucratic requirements.⁵⁵ The effort overall was seen as a huge innovation in how the city government operated, streamlining processes for data access to address immediate data needs. The work additionally revealed how MODA's role as intermediaries between city agencies and external resources helped position MODA as the right team for the data stewardship work involved in RDP's design and implementation.

For the Emergency Management's Recovery Team specifically, the dashboard has helped the city remain aware of activity in the city and the ways this activity affects residents. Though the team discussed discontinuing the dashboard as cases began to rise in the fall and winter, it continued producing the dashboard (with updates every Friday) to allow the city to understand changes in activity over time. The dashboard and reports were almost exclusively internal, with analysis feeding indirectly into later publications.⁵⁶ While it was eventually

⁵¹ Umberger, Melissa. 2023. "RE: [EXTERNAL] Open Data Policy Lab: NYC Recovery Data Partnership Case Study," February 10, 2023.

⁵² Goldberg and Umberger *supra* note 39.

⁵³ Schmoeker *supra* note 7.

⁵⁴ Bjornsson *supra* note 6.

⁵⁵ Goldberg and Umberger *supra* note 39.

⁵⁶ The city did share information with the Red Cross and the Salvation Army, who serve as the city's emergency response partners. Source: Goldberg and Umberger *supra* note 27.

discontinued as the City closed the partnership, staff reported that some of the data was used to inform locations of the City Cleanup Corps, with participated in some sanitation efforts during the pandemic.⁵⁷

As previously discussed, the NYC Department of City Planning considered the primary value of the mobility data to be informing mid- and long-term planning that will support the city's recovery. Little public documentation on it remains available on this work, though staff believed the data to be helpful for their work.

While reports from individual analysts are promising, the lack of public information on either agency's work makes it difficult to assess the extent to which the partnership has succeeded in its initial goals to reveal "how New Yorkers have been impacted by COVID-19" and "to inform programmatic and policy decisions for an equitable recovery."⁵⁸ Additional analysis will need to be done following the release of products to assess specific outcomes.

RISK AND MITIGATION STRATEGY

As reported by organizers and participants, the NYC Recovery Data Partnership had a robust approach to address risks. Each participating organization in the partnership conducted their own internal data protection and data responsibility procedures to ensure their compliance with local and national law. In addition, the City of New York, prior to forming agreements with these organizations, conducted reviews of their practices to ensure that partners abided by City standards for data collection and handling (e.g. having "opt-in" requirements and informing data subjects that their data may be shared). These standards and expectations were then outlined in a license agreement, which was made publicly available on the Partnership's website, ensuring some degree of transparency.⁵⁹

To guard against misuse of data and ensure there was appropriate buy-in within city agencies, all organizations seeking data had to make formal requests (that could be used to track decision-making around data) that were approved by agency leadership and legal counsel. Data supplies were not unlimited but often time- and purpose-bound.

While all these proactive efforts appear successful in mitigating risks, the research team notes that there has been limited transparency on individual requests or the results of them—which made it difficult for members of the public to provide input into data activity or to exercise meaningful oversight directly. It also may have also made the value of the NYC Recovery Data

⁵⁷ City Cleanup Corps. 2022. "City Cleanup Corps: A Foot in the Door to NYCHA - NYCHA Now." *NYCHA Now: Your NYCHA Employee Bulletin* (blog). May 18, 2022. <http://nychanow.nyc/city-cleanup-corps-a-foot-in-the-door-to-nycha/> ; Umberger *supra* note 52.

⁵⁸ Office of the Mayor *supra* note 9.

⁵⁹ City of New York Mayor's Office of Operations *supra* note 47.

Partnership difficult to understand for those outside it, potentially limiting support for the effort and undermining its overall sustainability. As of publication, most pages dedicated to the NYC Recovery Data Partnership have been removed from the open web and many resources are only available through the Internet Archive or preserved press releases. While the NYC Recovery Data Partnership was always intended to be a temporary initiative (and evaluations of it should recognize that fact), the “link rot” that has occurred in the months since its closure make external assessments of the effort difficult. Given changes in leadership across the City since the project launched, it also suggests the challenges of maintaining a data initiative across different administrations.

LESSONS LEARNED

Data users involved in the project had a few takeaways from their experience.

Enabling Conditions

- **Crisis Conditions:** As one might expect, one of the major motivating factors for the NYC Recovery Data Partnership was the crisis presented by the COVID-19 pandemic. The unprecedented nature of the outbreak provided impetus and justification to set up infrastructure that had previously not existed. City staff and partners felt pushed by the rapidly developing circumstances around them to act quickly to attempt to mitigate challenges.
- **Responsiveness of City Organizers:** One of the most surprising aspects of the project for its participants was the speed of the initiative. After requesting data, agency analysts received timely responses to their requests. The multi-layer reviews of use case applications tended to take about one week to conduct and complete.⁶⁰ This speed from the partnership’s leadership to review proposals, and share data meant agencies could begin work almost immediately instead of being bogged down by usual bureaucratic requirements.⁶¹
- **Responsiveness of Vendors:** Likewise, analysts involved in the partnership spoke highly about vendors who were responsive to their needs and concerns. Though one analyst reported that a provider was “opaque” and often took weeks to respond to questions about the data, the experience with all other providers was positive. Most providers engaged directly with analysts, created Slack channels to facilitate conversations with both providers and other institutional users. Discussions touched on applications of data, methodological best practices, findings from other users, and a variety of other topics. This Slack group’s contributions proved critical in ensuring

⁶⁰ Bjornsson *supra* note 6.

⁶¹ Goldberg and Umberger *supra* note 27.

the agencies, which did not have prior experience with mobility data, could use the data in a robust and meaningful manner.⁶²

- **Technical Capacity:** Finally, the team involved in the NYC Recovery Data Partnership attributed their success to the city's overall technical capabilities. As a large metropolitan area with significant obligations even outside crisis events, city agencies already had technical capacity to work with very large datasets on the end user side. While the team had to learn how to structure data for the Recovery Data Partnership in a way that facilitated data sharing, the existing baseline knowledge enabled early conversations and ensured they could process and analyze the data, even if they had not previously worked on mobility data specifically, without having to bring in a third-party consultant. Analysts expressed doubt that a smaller government without highly skilled workers could have achieved this work at the same scale.⁶³
- **Community of Data Analysts:** According to members of the Mayor's Office of Data Analytics, one major enabling condition for the Recovery Data Partnership was its existing community of City agency analysts via an internal community group called the Analytics Exchange (AnEx). These professionals, who often already knew one another, were a known user base for the Partnership that city leaders were able to focus their efforts on. In addition, while many analysts already knew one another through prior City data efforts, the context of COVID-19 encouraged many to collaborate. Analysts worked closely together to address questions and provide support as the need arose.⁶⁴

Challenges

- **Data Limitations:** Mobility data derived from smartphones tends to over-represent wealthy and younger populations more likely to own smartphone devices. While this data can still produce useful insights, analysts noted they needed to caveat research derived from this input and stress to policymakers what the asset could and could not represent.
- **Lack of Data Engineering Capacity:** Staff at the Partnership reported that, while they had robust analytics teams across City agencies, there was "relative lack of data engineering capacity, which [was] crucial to empowering those analysts with data that is easy to work with and understand." Partners provided agencies with different kinds of non-traditional datasets, including both the exhaust of normal activities without documentation and commercially available products with robust documentation. It was at times difficult to understand how to use these disparate assets in a productive

⁶² bid.

⁶³ Ibid.

⁶⁴ Bjornsson *supra* note 6.

fashion, though the Department of City Planning's Enterprise Data Management Team, the team that built the data engineering infrastructure for RDP, did eventually "create automated pipelines that connected data from partner to analyst, making each RDP dataset more usable and interoperable wherever possible" as well as "write robust data documentation."⁶⁵

- **Accuracy:** Unlike other institutions, the work produced by government agencies can have a very tangible and direct impact on the quality of life of ordinary people. As such, analysts reported how they needed to hold themselves to a higher standard of accuracy, going to great lengths to verify findings and identify gaps in the datasets. This need to guarantee quality work slowed down analysis and, per the analysts, slowed down the public release of products derived from partnership datasets.
- **Desire for Additional Funding:** Though the city had the benefit of a highly skilled workforce with some experience in data science prior to the the pandemic, agencies still faced significant resource constraints due to the fiscal crisis. Analysts expressed a desire for more funding to be set aside to support sophisticated data analysis across agencies. By making more resources available, interviewees argued that agencies would be more likely to accept and use large datasets in new and interesting ways.

⁶⁵ Ibid.



Photo by Colton Duke on Unsplash

CONCLUSION / NEXT STEPS

The NYC Recovery Data Partnership represented an innovative approach to data collaboration amid a crisis. Though there are various lessons to draw from the experiences of its participants, the partnership emphasizes the value of collaboration, not just across sectors but different city agencies. When organizations understand how others use data, they can learn from their experiences and research to inform their own actions. In New York and worldwide, experience in collaboration can set a foundation for future success. The institutional knowledge and precedence gained by the NYC Recovery Data Partnership can support efforts to launch data collaboratives in the future and provide impetus to similarly innovative efforts.⁶⁶

⁶⁶ While an earlier version of this case study was made available to interviewees and other city officials, city staff declined to review the final version of this case study pre-publication.



Open.Data.Policy.Lab

opendatapolicylab.org