COVID-19 Tech Taskforce
Workstream Name: Clinical Health Data

Members: AWS, Color Genomics, ESRI, Gates, Google, Ro, NAVIGO

Goals and opportunities

- Assist researchers at the CDC, academic institutions, and other public-health officials to capture clinical location data to assess resource capacity, effectiveness of policy interventions, community education, and disease surveillance.
- Gather patient information (including behaviors, symptoms, and risk factors) in order to create a model that can act as a high-value entry point.
- Assist health systems with de-identification and anonymization tasks.
- Understand and comply with relevant privacy and health regulation without impeding our effectiveness.
- Clearly define the problem we are trying to solve so we can coordinate and execute as quickly and effectively as possible.
- Leverage data to help prevent an unnecessary bottleneck on our medical resources.
- Clearly identify the symptoms of COVID-19 in particular geographies to promote effective testing and healthcare response.

Planned operation outcome

Identify and provide clinical datasets for use in developing forecasting models and informing disease surveillance.

Ideal operation outcomes [disclaimer: we don't know whether any/all of these are likely or their feasibility; they're meant to be illustrative of the types of analyses possible with broader clinical-data availability]

- Aggregate accurate information at federal, state, and county level.
- Develop forecast models predicting incidences and correlating them to key data points, like number of available hospital beds and resource needs.
- Maximize limited testing resources by ensuring that the right groups of individuals are tested. For instance, this could be determined based on the location data of individuals who have tested positive.
- Along with location-based information, determine the cause of the spread of COVID-19 and mitigate continued spread. For instance, consider creating a rating system indicating the danger of the spread of COVID-19 in certain communities.
- Track the different strains and expressions of the illness. There are multiple strains of varying severity in Europe.
- Develop contact tracing solutions to help reduce spread at the top of the funnel. This approach has been used successfully in other countries.

- Follow interoperability standards like FHIR to allow for easier collaboration.
- Understand comorbidity, including the symptoms displayed and how they manifest so healthcare can improve outcomes.
- Determine whether task force resources are best spent assisting CDC and other agencies with their own questions, or in posing our own.
- Reduce the likelihood of additional mutagenesis. This could include a review of the potential presence of the virus in the United States at an earlier stage.
- Provide treatments tailored to specific strains of COVID-19. Two key considerations:
  - (1) Maintain a supply chain that accurately allocates emerging medicines;
  - (2) Set clear guidelines and establish a decision tree reflecting the risk/reward of particular kinds of treatments in different geographies.
- Understand interagency / intergovernmental collaboration and information sharing.

Other known working groups / efforts
- NAVIGO has a public-facing assessment that could be used for screening, education, data creation/collection, and other ideas to help improve the interaction at testing centers. Prototype here:
  (b) (4)
- CTA's public link to telehealth companies questionnaire
- There is a substantial amount of interagency / intergovernmental work that is already happening; one of the challenges will be to ensure appropriate coordination among government entities and the task force.

Concerns

- Working effectively while complying with regulation, including HIPAA and PHI, and maintaining ethical data-sharing policies. Key considerations here will be understanding who is using the data (e.g., is it open source) and what level of precision will be provided.
- One meta-consideration is the government's tolerance for data aggregation / centralization. I.e., should we (1) gather large amounts of data and allow for good centralized governance to use that data to make decisions or (2) should we work to narrow the scope of questions now so we don't expend time sorting through information after we've gathered it?

Members: AWS, Camber, ESRI, Facebook, Google, Harvard School of Public Health, Microsoft, Mozilla, R4, SAP, Salesforce

Goals and opportunities

To assist researchers at the CDC, academic institutions, and other public-health officials, capture non-clinical location data to assess contact rates, social distancing, effectiveness of policy interventions, and disease surveillance.

Planned operation outcome

Create a secure, privacy respecting environment for researchers to aggregate and analyze data to forecast and model COVID-19 related location and clinical data.

Existing data

At the moment, there are no publicly available, or aggregated datasets generally available to researchers. Although some companies have aggregated data, the aggregated data are not made publicly available and, at the moment, are limited only to a set of academic researchers and epidemiologists. As of Tuesday, March 17, 2020, none of the members of the task force have agreed to share data through the task force without further definition of requirements.

Concerns

The primary concerns related to privacy relate to understanding the extent of consent at the point of data collection, considerations around deidentification and aggregation, guarding against re-identification, maintaining clear guidelines around use of data (including the roles of individuals accessing data, data storage/retention, and data reuse), and around transparency of work. Additionally, security and compliance (both technical and regulatory) present challenges for any data sharing effort.

Anticipated Deliverables

- A "Principles" document. We need a set of agreed-upon principles to help guide technical and implementation details. This will guide all subsequent deliverables for the task force.
- Assistance to entities around deidentification/anonymization efforts (provided sufficient regulatory / legal approval/authority)

COVID-19 Tech Taskforce
Workstream Name: Location Data

Members: Camber, ESRI, Facebook, Google, Harvard School of Public Health, Nielsen, Unacast, Uber, Veraset

Goals and opportunities
To assist researchers at the CDC, academic institutions, and other public-health officials, capture non-clinical location data to assess contact rates, social distancing, effectiveness of policy interventions, and disease surveillance.

Planned operation outcome
Identify and provide non-clinical geolocation datasets for use in developing forecasting models and informing disease surveillance and policy actions to prevent further spread of the disease. We should leverage existing data to make mathematically informed decisions regarding where to place resources.

Possible questions to be answered with our models

- Are social distancing policies being obeyed? Are they effective?
- Are people staying at home? How do we support them?
- What kind of specific messaging works and what doesn't to encourage desired outcomes?
- What is causing spread in one area, and how does it allow us to mitigate the risk of spread in other areas?
- What patterns are we seeing in transportation? Are people moving away from public transportation in favor of private means?
- How can we use data collected to ensure we maintain a supply chain that allows the necessary resources to be allocated to those communities that need them most (e.g., respirators, test kits, etc.)?
- How can we coordinate inbound requests to private industry for information to different companies from different agencies and branches of government?

Possible products

What are the deliverables that government needs from us? We proposed the following:

- Development of a data pipeline
- Development of coordinated dashboards
  - Ideally dashboards will be considered an authoritative source
  - Information could be gathered from WHO, CDC, and foreign governments

○ Information gathered could include: number of individuals tested, confirmed infected, still under investigation, school closures, unemployment, other business metrics, etc.

Existing data

- Mobile location data (Camber/Veraset/Unacast)
- Trip data (Uber)
- Dashboards (ESRI)
    - Data on hospital beds, number of beds (existing), staffed beds (supported), ICU beds and bed utilization rates.
    - https://coronavirus-resources.esri.com/
- Media behavior changes (Nielsen)
- Potential sociodemographic information and point of interest information

Other known working groups / efforts

Concerns

- Engaging federal, state, and county government so we don't lose time in coordinating
- Complying with regulation and establishing the appropriate data aggregation threshold
    - Clear messaging that data is aggregated and not personalized
    - Clear about our data aggregation threshold. This will be different based on size and density of community, and whether it's an open or closed dataset
- Privacy (see Hosting/Privacy workstream)
- Coordination
- Redundancy
- Sparsity of data / data quality
- Pipeline access/aggregation/hosting
- Update frequency
- Ethics/punitive use of data
- Ensuring the models we build are effective in future, and not models of past. We need to start predicting, or we won't be helpful in allowing gov to execute on plans and policies for current use case of prevention

Requests to the government

- Who do we coordinate with in government and relevant agencies (FEMA, Census, etc.)?
- What are the top questions that we can answer for government? This information will be key in helping us inform our approach and get started more quickly
- Data use agreements
  - Public access files, public utility files
- How should we coordinate with Federal, State, and local government most effectively?