

The #Data4COVID19 Review: What is non-traditional data? How was it used during COVID-19?

[The GovLab](#)¹ sought to evaluate how non-traditional data (NTD) was used to respond to the COVID-19 pandemic and lessons learned for future data-driven crisis management. This document includes the most prominent NTD use cases identified during this effort. Our intention is to raise awareness of the potential of specific types and sources of NTD during crisis management while also demonstrating its limitations. Learn more by visiting our [website](#) or reading the [full report](#).

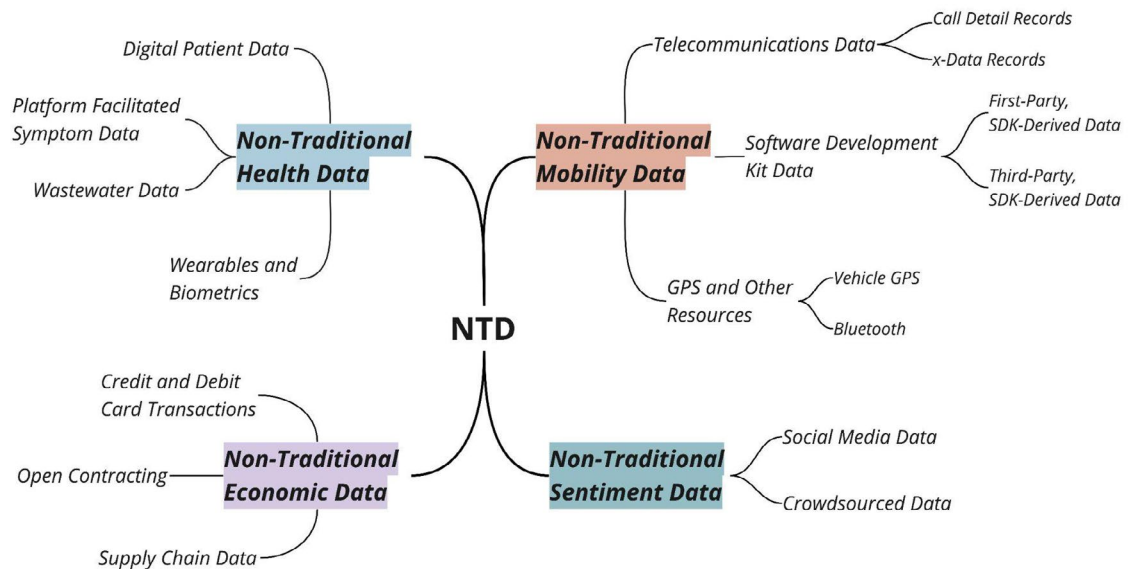
What we mean by NTD

[NTD](#) refers to data that is digitally captured, mediated or observed using new instrumentation mechanisms, often privately held and used for purposes unrelated to its initial collection. NTD is frequently used to complement traditional data. The use of NTD often allows decision-makers to increase the speed from collection to dissemination, scale, and quality of data. At the same time, given the limitations of NTD as it relates to representativeness, the use (and in some cases collection) of traditional data was needed to validate the insights.

	Traditional Data	NTD
What it includes	Data collected about a topic	Proxies for the data we intend to measure
Where it is located	Analog + digital systems	Digital systems
Data Type	“Small”, Structured	“Big”, Often Unstructured
Why it is generated	Data collected for a specific purpose	Data that has been repurposed for a new use case
Examples	Surveys, data collected by National Statistics Offices	Social media data, crowdsourced data, telecoms data

How NTD was used during COVID-19

Decision-makers around the world have called upon NTD to respond to different aspects of the COVID-19 pandemic. NTD was primarily used across four main areas of interest: health, mobility and geolocation, economics, and sentiment. Several sources of NTD were used within these four areas of interest throughout the different waves of COVID-19. The data sources that stood out most during our analysis along with key use cases are listed below.



Non-traditional health data describes the physical health of individuals or populations as it relates to COVID-19 diagnoses and risk factors. Non-traditional health data sources were primarily used to validate, supplement, or direct traditional COVID-19 testing data and understand how the virus impacted different population segments:



- **Digital patient data** refers to electronic health records and metrics of COVID-19 patients. While the use of patient data during a pandemic is not new and varied between countries, COVID-19 accelerated data collaborations by combining digital patient data from medical facilities, private sector organizations, and other sources. For example, [The National COVID Cohort Collaborative](#), a public-private partnership, was launched in September 2020 and has become one of the largest publicly available digital patient databases in the United States.
- **Platform facilitated symptom data** refers to the use of new survey instrumentation methods such as internet searches and mobile applications to gather information from the public about COVID-19 symptoms with the goal of accelerating research and finding new indicators of infection. One such example is the [ZOE](#) mobile application in the United Kingdom which used self-reported symptom surveys to identify early indicators of infection and other factors.
- **Wastewater data**—which is a combination of household sewage, industrial run-off, and, in some places, storm water—can contain trace amounts of the SARS-CoV-2 virus and detect COVID-19 among asymptomatic and untested individuals, allowing for earlier detection of outbreaks. In Hong Kong, [wastewater surveillance](#) efforts found the first evidence of the spread of the delta variant.
- **Wearables and biometrics** includes data about the human body collected using wearable technologies (e.g. smart watches), sensors, and other devices to identify individuals with physical characteristics that could indicate COVID-19 and monitor how the virus impacts the body. [Kinsa Health's smart thermometer](#) could detect atypical fevers and predict outbreaks across the United States.

Non-traditional mobility and geolocation data refers to the physical location of an individual or object in relation to one another or geography. Throughout the pandemic, public officials and health experts mainly used non-traditional mobility data sources to understand the spread of the virus and the impact of control measures:



- **Telecommunications data** refers to data provided by telecommunications companies that tracks individuals' locations and movements. There are two common types of telecommunications data: call detail records (records of each time a mobile phone connects to a network when sending or receiving a voice call or SMS) and x-data records (internet connection detail records generated when a device such as a cell phone or tablet connects to a mobile internet). In [the TELUS Data for Good initiative](#), the Canadian telecommunications company provided its data to the government to study the COVID-19 crisis.
- **Software development kit data** includes location data collected through smartphone application software using the device's hardware. Two types of software development kit data used during COVID-19 were First-party, SDK-derived data (data collected directly from a smartphone application without intermediary brokers) and Third-party, SDK-derived data (data collected from smartphone apps by many sources aggregated into one database). In the [Teralytics project](#) the Zurich-based mobility data company provided dashboards to health authorities, emergency responders, mobility providers, and transportation planners based on its collected third-party SDK-derived data.

- **Vehicle GPS and other resources** such as bluetooth were also used to track population movements during COVID-19. In April 2020, [Google and Apple](#) announced they would support health agencies by building cross-device Bluetooth-based contact tracing functionality into their devices. This functionality, known as [Exposure Notifications](#), was used by various health departments to develop [mobile applications](#) that would notify users when they came in contact with someone who had indicated they were sick with COVID-19.

Non-traditional economic data indicates the economic activity of individuals, groups, or organizations. Non-traditional economic data sources allowed policymakers to gain further insight into how businesses were reacting to lockdowns and how citizens' spending habits were changing as the virus and associated health policies evolved:



- **Credit and debit card transactions** generate data about where, when, and how people spend money as well as how much money they are spending. Statistical organizations have used these transactions as a proxy for economic health amid the pandemic. For example, The Bureau of Economic Analysis of the United States used daily card data from [Fiserv](#), a financial technology company, to measure the reduction in revenue of local businesses around the time of the pandemic.
- **Supply chain data** refers to all the data referring to all data about [supply chain processes](#) that inform the production and distribution of goods by analyzing trends and providing organizations with necessary insights to create an optimal supply plan. Using maritime traffic data collected via a global network of Automatic Identification System (AIS) receivers, a study from [the Marine Traffic Research Lab](#) investigates how the COVID-19 pandemic and the consequential social containment measures related to the shipping industry, which accounts alone for more than 80% of the world trade.
- **Open contracting** refers to using data from [public contracts](#) that are made available to the public. In Ecuador, the National Public Procurement Service launched a [public facing dashboard](#) of emergency procurement contracts during COVID-19.

Non-traditional sentiment data explains the attitudes and perceptions of individuals and groups regarding developments related to COVID-19. Non-traditional sentiment data sources were mainly used to understand the public's perceptions and attitude about pandemic response measures, the impact of misinformation and supplement traditional data sources such as surveys:



- **Social media data** gathered from social media platforms like Twitter, Facebook, and Instagram, can provide insights into how people are feeling about different aspects of the pandemic and the response to it. For example, [Civicytics](#), a South American initiative launched by IDB Group, used data from social media and applied AI to understand how people were feeling about the pandemic and the response to it.
- **Crowdsourced data** through mobile applications and online surveys were used to understand how people felt about specific COVID-19 related issues and policies. This data was used to understand the pandemic at a granular level and identify patterns or trends that would be otherwise difficult to discern. [CoMix](#), an initiative developed by Hasselt University and University of Antwerp and funded by the European Commission, combined crowdsourcing tools with data analytics to quickly gather public sentiment data from across the European Union.