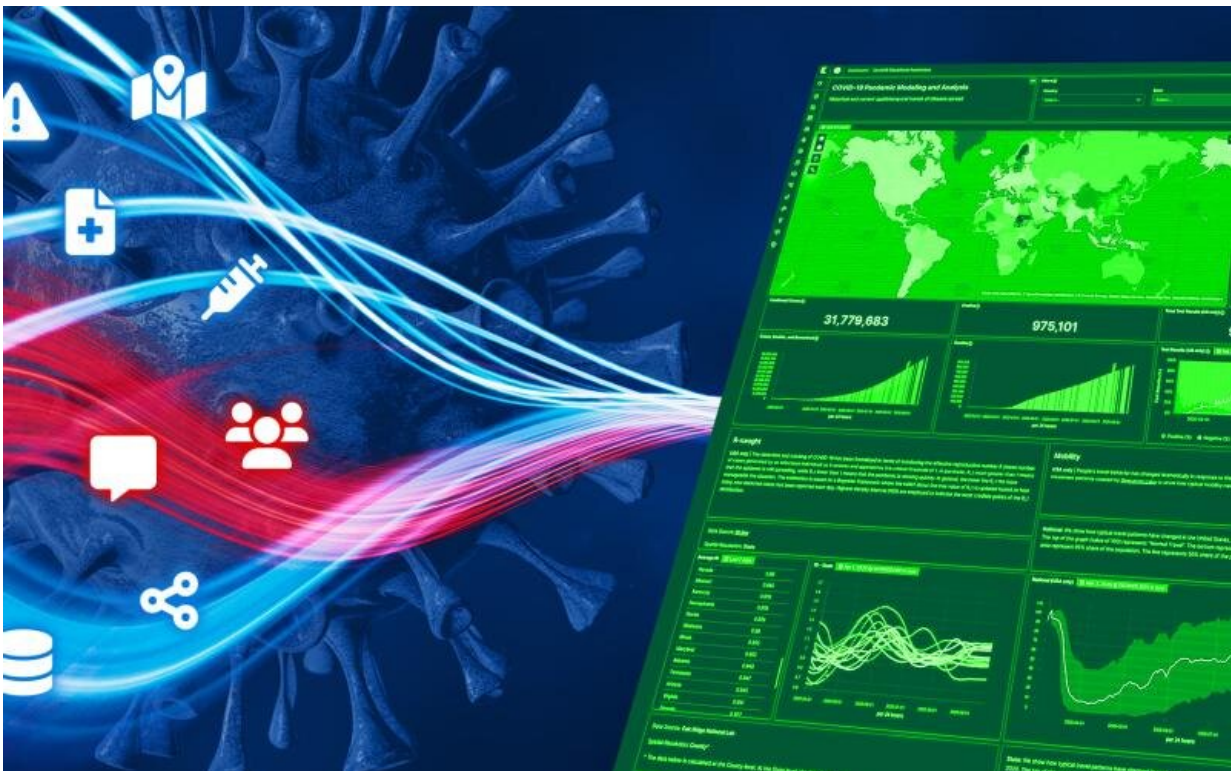# Data science teams analyzed COVID data for early pandemic response

July 27 2022, by Liz Neunsinger

Data from different sources are joined on platforms created by ORNL researchers to offer better information for decision makers. Credit: Nathan Armistead/ORNL, U.S. Dept. of Energy

When the COVID-19 pandemic stunned the world in 2020, researchers at the Department of Energy's Oak Ridge National Laboratory wondered

how they could extend their support and help. The health crisis demanded medical professionals to work on the front lines caring for ill patients; but, what about those not in medicine? ORNL researchers who focus on human security quickly turned to what they do best: They crunched data to provide actionable information for the COVID-19 response.

In two separate projects, teams of scientists combined various types of publicly available information to create dashboards of COVID-19 data accessible to authorities to make informed decisions. In both projects, the teams found disorganized and unstructured data that varied greatly in how demographics, incidents and severity of infection, response to pandemic restrictions, and infection recovery were captured and reported.

"One thing that surprised me was the number of data sets we received," said Gautam Thakur, Location Intelligence group leader at ORNL. Thakur and his team stood up a platform and created an online dashboard for monitoring COVID-19 progression to map geospatial infrastructure, such as where hotels, stadiums, and convention centers were being used as medical facilities; data about patient infection and recovery data; or local policy data. The COVID-19 Joint Pandemic Modeling and Analysis Platform provided real-time awareness of the changing situation down to the county level across the nation. Platform construction was published in *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19*.

The team collected data from authoritative sources, such as the Centers for Disease Control and Prevention and John Hopkins Coronavirus Resource Center; private corporations; and non-authoritative sources, such as social media and public forums. The platform structured data to allow for complex analyses across locations and time intervals.

The COVID-19 platform was used in ORNL's backyard: Knoxville, Tennessee. It predicted outcomes and gave authorities more fidelity on which decisions would benefit the public, whether they were evaluating how creating a new place to accommodate patients could reduce new cases to understanding how vaccine disinformation impacts vaccine uptake.

The platform automates data ingestion before processing it through machine learning, deep learning analysis or statistical modeling using high performance computing at ORNL. Users could then access and use its applications to visualize the findings presented in a way comprehensible to a non-technical audience.

How his team came together to build the platform is what Thakur remembers about those first few weeks working during the pandemic. "We can make a tangible impact on problems in a very short amount of time." Over the grueling months when data was vital to a fast-changing situation, their platform hosted hundreds of people 24 hours a day, seven days per week, with 99.9% availability.

Teamwork was a theme for another project aimed at helping authorities understand COVID-19 infection trends. Robert Stewart, an ORNL senior scientist in Geospatial Artificial Intelligence, and his team built a data driven seven-day model based on new infection rates, one of the few types of data reliably available in the early days of the pandemic. The model assisted government agencies in understanding how the virus had moved through the country over the previous week and anticipated changes in the upcoming week.

"It was very rewarding to see a data-driven model have such operational success, especially in the rapidly changing data environment we faced in those first weeks and months," said Stewart. "It was a tremendous effort by the entire team to design and apply data science methods that

produced reliable COVID maps for the country in such a short time. Those maps were updated and made available to the federal government on a weekly basis." A summary of this model is published in a Department of Energy Report on Rapid R&D Solutions to the COVID-19 Crisis.

Early on, moving beyond new infection rate data was problematic. Each day, the team scraped data from websites across the U.S. to get better details about recovery, deaths and demographics. They found a lot of variability in the way the data was reported from location to location.

Some data was consistently available, such as COVID-19 cases and deaths. Other demographic data differed across regions. Daily reporting from hospitals was inconsistent, was offered at irregular intervals and was often incomplete.

Geography researcher Alex Sorokine, an ORNL researcher on this project, seeks to improve the application of data science to emergency response situations in two areas. First, he recommends standardization: encouraging health providers to produce data in consumable formats would help data scientists better capture the relevant variables. Second, Sorokine urges further expansion of natural language processing using artificial intelligence and machine learning algorithms to distinguish how data is collected. Humans on the front lines of a pandemic may use different words or categories for similar pieces of information, and a computer can detect which types of information are similar.

The disparity of information proved to be the hardest part of the project. ORNL's Jason Kaufman, a geospatial data curator, joined this effort from the onset and was ready to get answers for authorities needing information. However, the lack of consistent data meant he manually configured data points each day. Through the strenuous time, he never thought about stopping.

"That wasn't an option," Kaufman said. "We knew this was data that didn't exist anywhere else. We had to do this so someone could use it."

Kaufman is ready to support his team the next time an emergency requires data science. As a person who recognizes the puzzle of data, he is ready to solve the next one. Building a dataset that no one else has ever created before, he said, is about as cutting edge as you can get for data science.

Provided by Oak Ridge National Laboratory