

Covid Engineering Professor Data Files WSJ 051120 print

<https://www.wsj.com/articles/how-a-johns-hopkins-professor-and-her-chinese-students-tracked-coronavirus-11589016603?mod=searchresults&page=1&pos=2>

How a Johns Hopkins Professor and Her Chinese Students Tracked Coronavirus

Imperfect, yes, but real-time pandemic tracing is a first; '4,000 dots on my dashboard'

By Jon Hilsenrath

Updated May 9, 2020 4:35 pm ET

Lauren Gardner, an associate engineering professor at Johns Hopkins University, was sitting in a school coffee shop with two graduate students in January, chatting about their work on measles, Dengue fever and the Zika virus, when conversation turned to an emerging [coronavirus](#) in China.

The students, both from China, wanted to track it, and Dr. Gardner, who researches how transportation systems propagate the spread of disease, was game. They built a website in a day, a [Covid-19 dashboard](#), and after launching it on Jan. 22, found themselves at the center of a historic first: tracking a deadly, fast-moving pandemic as it spread globally.

The website became a far bigger hit than anyone anticipated. With swelling bright red dots and grim tallies of death across nations, it has become an indispensable resource for anybody following the pandemic, a centerpiece in situation rooms around the world—including the operation center at the U.S. Department Health and Human Services—and a regular feature on television newscasts.

In March the website that hosts the online map, [arcgis.com](#), generated nearly 1 billion visits, according to web tracker SimilarWeb. That was more than LinkedIn, CNN or eBay, and it was among the 100 most visited sites in the world, according to SimilarWeb data. President Trump stood before a chart built in part with data from the Johns Hopkins tracker that month when announcing a call for social distancing.

The U.S. government hasn't produced anything comparable. Others have followed with similar sites, including news organizations, the [World Health Organization](#) and a nonprofit run by software developers called [Worldometers](#). A U.S. [Centers for Disease Control and Prevention website](#) includes similar U.S. data, but doesn't update as frequently.

"While case counts are not perfect, they are invaluable when compared with having nothing," said David Finnoff, a University of Wyoming economics professor who studies pathogens and pandemics. "I look at the website data, or media applications of it, many times per day."

Dr. Gardner's experience is a window into the challenges of chasing a rapidly spreading virus and understanding its health implications. Sparse testing, asymptomatic cases and disparate government methods for reporting infections and deaths have made it hard to know the scope of the disease. In some places, moreover, officials who produce data that feed the website may have incentives to misreport numbers.

Dr. Gardner thinks coronavirus cases and Covid-related deaths have been undercounted. She's also near completing a study that finds social distancing helped

slow the spread of the virus, just as many states, businesses and households tire of distancing policies and the economic damage they cause.

“Absolutely it works. We have very clear evidence,” Dr. Gardner said of social distancing in an interview. “We have observed data to quantify social distancing and observed case data to quantify the outbreak growth rate and can show there is definitely a strong correlation between the two.”

The 35-year-old Texan has an engineering Ph.D. from the University of Texas at Austin. She arrived at Johns Hopkins last year after spending several years as a civil engineering lecturer at the University of New South Wales in Sydney; she and her husband named their pet dog Tim Tam after an [Australian snack](#).

The project began in almost slapdash fashion. One of Dr. Gardner’s students, Ensheng Dong, was worried about family back home in Shanxi province, north of Wuhan. A geography and mapping specialist who has studied and worked in the U.S. since 2012, he spent long hours inputting data by hand while also taking classes.

They started with a couple of websites, news reports and their [Twitter](#) feeds for data. Because the disease was centered in China, a primary source was DXY.cn, an online community for Chinese medical professionals that tracked official coronavirus counts locally.

“We were thinking this would be really cool if maybe like dozens or a hundred researchers ever want to use it,” Dr. Gardner said. Mr. Dong also wanted data for his dissertation. “We pulled it together that night.”

The site became more sophisticated and automated as the disease grew and people turned to it for updates. The team started scraping official figures from U.S. cities and states, in addition to official websites from countries, provinces, cities and other government entities around the world. Dr. Gardner says there are roughly 7,000 data points in all to track from dozens of sources, including several data-aggregation sites.

The team, which expanded and drew in resources from elsewhere at Hopkins, now spends most of its time staying on top of data pouring in. That includes an “anomaly detection system” to spot quirky inputs, she said. It also shifted from campus to working remotely.

Many government officials, in turn, relied on it to trace where the virus spread.

“At the very beginning, it was far and away the best information that we had about what was going on around the region and the U.S. in real time,” said Max Reiss, a spokesman for Connecticut Gov. Ned Lamont.

The map was on a big screen when Mr. Lamont and state officials crowded around a table in an emergency operations center on March 10 to discuss declaring civil-preparedness and public-health emergencies.

The federal government’s coronavirus task force relied on modeling from Seattle’s University of Washington Institute for Health Metrics and Evaluation, which in turn relied on data on deaths gathered in the Hopkins Covid-19 dashboard, according to a spokeswoman for the IHME. On many days, coronavirus task force meetings in the White House Situation Room began with Dr. Deborah Birx, the White House coronavirus

response coordinator, and Dr. Anthony Fauci, the government's top infectious-disease official, reviewing the IHME models.

Behind the maps is a complicated data supply chain replete with challenges that come from gathering and packaging raw information from thousands of sources around the world.

U.S. counts reflect those challenges.

Testing varies across the nation. New York has conducted 6,464 tests for every 100,000 individuals in the state, according to Hopkins. That compares with 1,526 per 100,000 in Arizona. With testing limited and varied, and many infected people who never show symptoms, it is hard to know the true spread of the virus.

Tabulating deaths is tricky too. Some states count probable deaths for cases where there weren't test results available but where the deceased had symptoms of the disease, in line with the Centers for Disease Control and Prevention's guidance. Other states don't yet include probable case data in the numbers they post online.

Deciphering cause of death comes down to individual judgments made by a wide tapestry of doctors, medical examiners and coroners.

Methods differ even within states. New York state only reports test-confirmed Covid-19 deaths in its official online tally. New York City, on the other hand, started adding probable cases to city-specific data in mid-April, conforming with CDC guidelines. Dr. Gardner's team manually added [thousands of probable deaths](#) to the dashboard as a result.

Dr. Gardner said it isn't her team's role to adjudicate these differences. Where authorities report probable cases, she'll include it.

"How am I supposed to figure out how many probable deaths there are in all 4,000 dots on my dashboard?" Dr. Gardner says. "It is an impossible thing for me to do. It is hard enough to just to pull in the data that's been reported."

The challenge is global.

In early April, France raised its official death toll by more than 40% after picking up cases that had been undetected in nursing homes.

The Hopkins site also digested revisions from China, where the [possibility of significant undercounting](#) has aroused suspicions among epidemiologists, U.S. intelligence sources and Wuhan residents.

Dr. Gardner said she believes that China, like other countries, including the U.S., has undercounted cases. She said she couldn't say if this is due to "malicious intent" or just the challenge authorities everywhere face keeping pace with a fast-moving, new disease.

In past disease epidemics, countries tended to delay reporting until a problem became certain, because notifying others came with economic costs, such as reduced travel and trade, says Ramanan Laxminarayan, a researcher at Princeton University. Early warning systems, he and others said, are key to prevention.

“In some ways, it does seem like a task the CDC could do,” Dr. Gardner said of building a U.S. tracking system. “I would say that they’ve taken on a lot of other tasks and focused on other things that are very complementary.”

It wasn’t clear at first the Johns Hopkins system would last more than a couple of weeks. The virus itself could have flamed out quickly. Instead, the tally Friday night showed more than 3.9 million infection cases and more than 274,000 deaths world-wide.

“There was this clear, huge gap for providing data publicly,” Dr. Gardner said. “This is the first time that the world has been able to watch in real time as a pandemic unfolded.”

—*Stephanie Armour contributed to this article.*

Write to Jon Hilsenrath at jon.hilsenrath@wsj.com and Jon Kamp at jon.kamp@wsj.com

https://www.wsj.com/articles/notable-quotable-nate-silver-11589139737?mod=article_inline