Coronavirus sequence trove sparks frustration

Users of GISAID complain of opaque decisions and onerous restrictions

By Meredith Wadman

In December 2020, software developer Angie Hinrichs at the University of California, Santa Cruz (UCSC), applied for access to a labor-saving data feed from GISAID, a nonprofit database of viral sequences including those of the pandemic coronavirus, SARS-CoV-2. She wanted GISAID's data so she could display mutations on UCSC's coronavirus Genome Browser. That tool ties any position in the virus' nearly 30,000-letter genome to other scientific information, much as Google Maps shows gas stations and restaurants near addresses.

With more than 700,000 genomes from more than 160 countries, GISAID is by far the world's largest database of SARS-CoV-2 sequences. Access to the free, nonprofit repository has become vital to Hinrichs and thousands of other scientists and public health agencies tracking the virus' alarmingly rapid evolution.

But instead of getting a direct data feed, Hinrichs lost her existing access to two conveniently packaged GISAID files that are the next best thing. She emailed GISAID repeatedly pleading for restored access, but hasn't gotten it. Since December, she has had to download GISAID's sequences 10,000 at a time, with no access to most of the metadata unless she looks at each of the 10,000 sequences individually. As a result, she says, "My [phylogenetic] trees that use GISAID data are falling behind."

Hinrichs's experience is not unique. A dozen scientists spoke with Science raising complaints about their interactions with GISAID. They reported an opaque process of gaining access, unexplained interruptions once access was won, and phone harassment or threatening legal letters when they got on the wrong side of GISAID's strict rules against re-sharing data. Many scientists who voiced criticisms declined to be identified for fear of losing GISAID access. They say that even as they race to study coronavirus evolution, they are walking on eggshells around their chief data supplier.

"I am so tired of being scared all the time, of being terrified that if I take a step wrong I will lose access to the data that I base my research on," says one scientist who declines to be identified. "[GISAID] has that sword hanging over any scientist that works on SARS-CoV-2."

In a statement, GISAID said, "Any individual who registers with GISAID and agrees to the GISAID terms of use will be granted access credentials. ... On rare occasions, GISAID has found it necessary to temporarily suspend access credentials to protect the GISAID sharing mechanism."

Both fans and critics emphasize that GISAID has provided an invaluable service during the pandemic, gathering many more coronavirus sequences than open-access databases like the United States's GenBank (see graphic, p. 1087). Even critics note that data are much easier to upload to GISAID than to open-access repositories, and that GISAID speedily curates sequences.

"GISAID has done an amazing job. They really have revolutionized access to all these data," says David Haussler, a computational biologist at UCSC who is Hinrichs's boss. "We really, really want give them credit for what they have accomplished."

Many scientists trace what they view as a secretive, controlling organizational culture to GISAID's co-creator and head, former Time Warner studio executive Peter Bogner. GISAID "has a personality behind it that is fiercely protective of the organization [and] very insulted if somebody else ... is praised for SARS-CoV-2 data," Hinrichs says.

Bogner has said he invested several million dollars to launch GISAID in 2008. Its goal was to open up access to then-restricted avian flu sequences, and to protect scientists in non-Western countries against having their data scooped for publication or profit by requiring users to credit and try to collaborate with depositors (Science, 25 August 2006, p. 1026 and 16 February 2007, p. 923).

GISAID, which stands for the Global Initiative on Sharing All Influenza Data, is today supported by private donors, governments, and nonprofits and is based in Germany; it says it remains "independent of government and corporate interests."

In its statement to Science, GISAID said scientists deposit to its database because they "are confident that their rights will be protected." Without GISAID, "We would now be in real trouble, because it's been successful in building confidence in SARS-CoV-2 genomic data sharing in countries around the world," says GISAID co-founder Nancy Cox, now retired from the U.S. Centers for Disease Control and Prevention.
But critics complain about GISAID’s constraints on access, chief among them its prohibition on reshar- ing of its data. Its agreement for access to the direct data feed also requires applicants to use only GISAID data in their websites and tools, as well as only GISAID-approved strain names.

Other scientists say the access process itself is opaque. Brooks Miner, an evolutionary ecologist at Ithaca College, contacted GISAID on 2 February hoping to get a data feed for a lay-friendly website mapping the frequency of coronavirus variants. He got a phone call with instructions from a man who refused to identify himself except as “a GISAID representative” and whose identity he still does not know. “I started calling him Mr. GISAID,” Miner says. (GISAID said it does not have a policy of not identifying its representatives.)

Puzzled, Miner contacted other GISAID users and found they lived in fear of losing access. “I realized people doing phenomenal cutting-edge science carry this fear that their career could be ruined on a whim by this faceless organization,” Miner says.

Miner was granted GISAID access last week but says he fears losing it because of his criticisms. “I’m speaking out anyway because I believe the way GISAID operates is flawed,” he says.

Other scientists say they have received threatening letters from GISAID lawyers. Early in the pandemic, Hinrichs pulled GISAID data from another organization, Nextstrain, and mistakenly failed to credit GISAID, prompting what she calls an “ominous” missive from a law firm directed to Haussler. “This was a new experience for us,” Hinrichs says. “We are used to speaking with scientists, not hearing from lawyers.” She added GISAID credits to the browser.

In its statement, GISAID said, “GISAID has never found it necessary to commence a legal action against a participant. ... We typically are able to come to a speedy and amicable resolution of any issues.” GISAID says it has revoked access for only one user in the past year, because they “would not abide by GISAID’s terms of use.”

Some scientists say they have gotten phone calls lecturing them on the virtues of GISAID and the flaws of public-access databases. “GISAID sees every sequence submitted to GenBank as a battle lost,” another scientist says.

Kelly Oakeson, chief sequencing scientist at the Utah Public Health Laboratory, which relies on GISAID data to track coronavirus variants in his state, recalls Bogner phoning him last year for a technical matter and then urging him not to deposit sequences in GenBank. He “really wanted to know … ‘What possible good could come of that? You’ve got it in one place, why do you need it in both?’”

GISAID denies disparaging GenBank or discouraging users from depositing in it or other open-access databases.

In January, scientists pushed back in an open letter urging scientists to deposit sequences in GenBank, the European Nucleotide Archive (ENA), and Japan’s DDBJ, open-access databases that allow users to access sequences anonymously and share data freely. “The ideal setup is completely open access,” to speed research, says signatory Guy Cochrane, head of the ENA. “Having a limited group controlling [access] would never be a good thing.”

GISAID countered in its statement that the letter “effectively calls for data to be shared anonymously and without any protection for the data contributors.”

Some users say they have only had good experiences with GISAID. “I’ve gotten much more support from GISAID than from any government agency,” says Jeremy Kamil, a virologist at Louisiana State University Health, Shreveport, and senior author on a recent preprint that identified seven new SARS-CoV-2 variants in the United States. He says he finds GISAID’s global, 24/7 staff responsive and helpful.

But others see much room for improvement. They want a right of appeal if they lose GISAID access and a transparent view of how GISAID is governed. They would like to open a conversation about ways GISAID might relax its data-sharing requirements during the pandemic, without risking their access by raising the subject.

Miner would also like to see a less territorial approach: “Aren’t we just trying to do good work that’s helpful in the pandemic?”

### Data dominance

By 4 March, GISAID had amassed nearly 700,000 sequences of the pandemic coronavirus from 160 countries—far more than any other database.

In January, scientists pushed back in an open letter urging scientists to deposit sequences in GenBank, the European Nucleotide Archive (ENA), and Japan’s DDBJ, open-access databases that allow users to access sequences anonymously and share data freely. “The ideal setup is completely open access,” to speed research, says signatory Guy Cochrane, head of the ENA. “Having a limited group controlling [access] would never be a good thing.”

GISAID countered in its statement that the letter “effectively calls for data to be shared anonymously and without any protection for the data contributors.”

Some users say they have only had good experiences with GISAID. “I’ve gotten much more support from GISAID than from any government agency,” says Jeremy Kamil, a virologist at Louisiana State University Health, Shreveport, and senior author on a recent preprint that identified seven new SARS-CoV-2 variants in the United States. He says he finds GISAID’s global, 24/7 staff responsive and helpful.

But others see much room for improvement. They want a right of appeal if they lose GISAID access and a transparent view of how GISAID is governed. They would like to open a conversation about ways GISAID might relax its data-sharing requirements during the pandemic, without risking their access by raising the subject.

Miner would also like to see a less territorial approach: “Aren’t we just trying to do good work that’s helpful in the pandemic?”

### COVID-19

Rich countries have ordered billions of doses more than needed for their populations

By Jon Cohen and Kai Kupferschmidt

Last week, U.S. President Joe Biden shaved 2 months off earlier projections and promised COVID-19 vaccine doses would be available for all U.S. adults by the end of May. That’s welcome news for the United States, which has had more COVID-19 cases and deaths than any other country. It could also be good news for nations that may benefit from the more than 1 billion extra doses the federal government has ordered.

Like three dozen other countries, the United States contracted with multiple vaccine companies for several times the number of doses needed to cover its population. No one knew at the time which, if any, of the candidate vaccines would work or when they might prove safe and effective. But by now, most of the purchased vaccines appear to offer solid protection—which means many countries will receive far more vaccine than they need. The excess doses the United States alone may have by July would vaccinate at least 200 million people. Over the next year or two, U.S. surplus doses and those from other countries could add up to enough to immunize everyone in the many poorer nations that lack any secured COVID-19 vaccine (see table, p. 1088).

No country yet has warehouses filled with unneeded COVID-19 vaccines—many hard-hit European nations may not have surpluses until the end of the year. But global health leaders stress that the time has come to figure out how to share those committed but extra doses. “It would be hugely beneficial to lay out a road map of 2021,” says Jeremy Farrar, who heads the Wellcome Trust biomedical research charity. “Planning is critical.”

Last summer, the World Health Organization (WHO) and other groups created the COVID-19 Vaccines Global Access (COVAX) Facility to address expected vaccine inequi-
Coronavirus sequence trove sparks frustration
Meredith Wadman

Science 371 (6534), 1086-1087.
DOI: 10.1126/science.371.6534.1086

ARTICLE TOOLS http://science.sciencemag.org/content/371/6534/1086

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service