

## Data availability, 21 March 2023

### **GISAID comments on the speculations surrounding data availability**

Recent discussions surrounding early hCoV-19 sequence data associated with a manuscript submitted by the researchers at the China CDC have been clouded by speculation regarding the status of the data. Here we clarify the current status of that data, and take this opportunity to reiterate the general data sharing policies at GISAID.

Data contributors frequently update their records with the help of GISAID's data curation team. When this occurs, released records will become temporarily invisible. Doing so recognizes the inherent need to balance the urgency of sharing genomic and associated metadata and its quality to aid public health decision making. GISAID does not, on its own accord, unrelease data. This decision rests solely with data contributors. See GISAID [data release policy](#).

#### *Background:*

For nearly 15 years, the GISAID data science initiative has been an essential contributor to global health security by enabling the rapid sharing of genomic and associated metadata during major public health emergency situations. The initiative has facilitated public health responses to outbreaks and the development of lifesaving countermeasures.

Thousands of data contributors from 215 nations and territories trust GISAID's transparent and equitable data sharing mechanism which incentivizes them to make available their data on currently circulating viruses in near real-time.

Among the many global contributors to GISAID are researchers from the China CDC. They have established a track record of promptly sharing their data during outbreaks with pandemic potential. For example:

In 2013, during the highly-pathogenic [H7N9 avian influenza](#) outbreak, they were credited in a [Nature Editorial](#) for "... rapid response ... and its early openness in the reporting and sharing of data [via GISAID]".

In 2020, in less than 48 hours from obtaining the first sequences that identified an unknown betacoronavirus from patients in Wuhan, the China CDC made the *first* whole-genome sequences available to the world via GISAID shortly after midnight on 10 January 2020 UTC, which enabled the development of lifesaving countermeasures to COVID-19 at unprecedented speed, including the first vaccines ([Polack et al N Engl J Med](#) 2020), and the first diagnostic tests to detect SARS-CoV-2 ([Bohn et al Clin Chem Lab Med](#) 2020) ([Carter et al ACS Cent. Sci.](#) 2020).

As China eased and eventually ended its lockdown policy in December 2022, researchers from 30 regions across China ramped up their genomic surveillance efforts and reported their latest hCoV-19 data through GISAID. Because some of these data were made available within just a few days from sample collection, it afforded the world a [transparent picture](#) of circulating variants across China and thus allayed fears that newly emerging variants of interest or concern might go undetected.

#### *Recent Discussions:*

The China CDC submitted additional sequences to GISAID that were obtained early in the pandemic, including raw data. Unlike current data from viruses that are circulating and evolving, these new data originated from environmental samples collected more than three years ago.

It is GISAID's understanding that after the China CDC submitted a first draft of their analysis to a major scientific journal, a first peer-review process yielded the request for more and improved data. The China CDC complied and provided the reviewers with improved and additional sequence data as part of a manuscript currently under review.

At the same time, some GISAID users accessed and downloaded an incomplete portion of these data. GISAID's terms of use explicitly permit users to collaborate with other GISAID users on analyses of this data.

However, if GISAID users were to publish an analysis of this un-published data before the data generators' own publication is released (*especially*, if they had knowledge the data generators submitted their own manuscript for publication), such an act would amount to scooping. Not only could this "distort[] scientific progress[,] it could negatively impact the quality of scientific work.

See Pearson, H. [It's a scoop!](#). *Nature* **426**, 222–223 (2003). Doing so would also run afoul to GISAID's Database Access Agreement, which calls on the user to make best efforts to collaborate with the data generators and involve them in such analyses and further research using such data.

Unfortunately, GISAID learned that select users published an analysis report in direct contravention of the terms they agreed to as a condition to accessing the data, and despite having knowledge that the data generators are undergoing peer review assessment of their own publication. When GISAID sought confirmation from the data generators whether best efforts to collaborate have been made in this case, GISAID was advised that a group of researchers contacted the data generators to communicate only their intent to *publish* their analysis of the generators' data. As such, the best efforts requirement has not been met. GISAID communicated the above-stated observations to these users on 14 March 2023 and provided them with the opportunity to comment, but did not receive a reply as of this writing.

GISAID's goal is to incentivize timely and transparent data sharing by providing a trusted place for data contributors to see their rights respected. It should be apparent to everyone that the data generators are the ones most familiar with the details surrounding their submitted data and the context in which it was collected.

This does not preclude the ability for other researchers to discuss analyses with peers or respond with publication of their own analysis once the data generators have been afforded the opportunity to have their analysis undergo a rigorous scientific peer review for purposes of publication.

GISAID is confident that the peer-review by the scientific journal will be expedited to the extent possible. GISAID strongly encourages that the complete and updated dataset will be made available as soon as possible to all GISAID users.

Premature discussions of the scientific data in the media risks eroding the public's confidence in scientific research.

[back to: Statements & Clarifications](#)