

From: Trevor Bedford [REDACTED]@bedford.io
Subject: Re: data packages made available for download
Date: October 27, 2025 at 9:59 PM
To: GISAID Secretariat [REDACTED]@gisaid.org, [REDACTED]@gisaid.org
Cc: Songer, Michael [REDACTED] GISAID Compliance [REDACTED]@gisaid.org, [REDACTED]@gisaid.org



Dear GISAID Secretariat and dear Peter,

Thank you for writing. I appreciate the dialog (though I would note again I would prefer a phone call as it would be more friendly and efficient).

First off, I deny that our analyses that are shared on nextstrain.org violate the terms of GISAID's Database Access Agreement. One immediate point here: we've operated the site in basically its current form since 2020, the only change being going from a single subsampled tree in early 2020 to different URLs with different subsamplings (for region, for timespan and for rooting, as discussed). It's the same basic analysis just subsampled in different ways. Furthermore, we've never heard anything about violation in the past despite it being clear that your team has reviewed our site. The clearest immediate example is when I had a Zoom call with Sebastian (cc'ed here) organized by Ron Fouchier and Marie-Paule Kieny on June 15, 2021. Sebastian brought up specific changes that GISAID would like. These were:

- 'hCov-19/' must remain as part of the strain name
- Authors in tooltip must include all authors
- Remove "DOWNLOAD" for GISAID data

I really appreciated the direct engagement from Sebastian (noting that I've since reached out repeatedly over past years for this sort of phone call to discuss specific requests from GISAID). The requested changes were immediately made. At this time in 2021 we had the same set of faceted trees available for different regions. Clearly we were not viewed as in violation of DAA terms in 2021 and nothing in the site has materially changed since then in terms of core functionality.

This said, I'd like to focus on moving forward and keeping genomic surveillance operating. I understand that you will not agree to flat file access for Nextstrain like you provide for LANL and CoV-Spectrum (though I would appreciate clarification on your distinction between these analysis websites and our own, and why these other services continue to have flat file access). In this case, I would like to proceed with your generous offer to have GISAID produce subsampled Auspice JSONs to host on nextstrain.org.

Once you provide GISAID with all necessary criteria for set selection, subsampling and target size for each tree. GISAID technical staff can gladly look into the modalities of enabling the 56 trees you are making reference to.

In this case, we can make a group account for GISAID that would live at <https://nextstrain.org/groups/gisaid>. This would work similarly to other groups where you can directly upload JSON files to our S3 bucket and these JSON files will then be immediately available as visualizations.

You can see an example of this with Africa CDC analyses like <https://nextstrain.org/groups/africa-cdc/ncov/algeria>. Here, this URL surfaces that these workflows were not run by the Nextstrain team and it gives direct attribution of the workflow maintainers. I would picture a similar setup for workflows run by GISAID on EpiCoV data. Auspice JSONs posted to <https://nextstrain.org/groups/gisaid> in this fashion would maintain restrictions like lack of metadata download to comply with the DAA.

We have documentation on Nextstrain Groups here <https://docs.nextstrain.org/en/latest/learn/groups/index.html>. We'd be very happy to work with your technical staff to get a groups account set up and to help delineate subsampling criteria.

I'd like to maintain the bright line between Nextstrain maintained analyses and analyses conducted by partners that are shared through nextstrain.org, which this scenario provides.

Thank you for your continued engagement,
- Trevor

On Oct 16, 2025, at 12:31 PM, GISAID Secretariat [REDACTED]@gisaid.org> wrote:

Dear Mr. Bedford:

Thank you for your recent correspondence. We have reviewed your request and determined that providing data in a format other than that proposed to you (i.e., an Augur output JSON file with the relevant subsample for your global tree of 'Genomic epidemiology of novel coronavirus (nCoV)'), is not workable given the terms of use for data in GISAID.

Your response, however, indicates that you are using the data received from GISAID for other non-approved purposes, and have "re-published" this data in various forms. GISAID authorized and maintained a data feed only for a specific purpose, i.e., for the updates of the 'nextregion' package. But since it was no longer being downloaded in a meaningful manner, it is no longer being produced. Use of data provided by GISAID for other purposes is improper and violates the terms of GISAID's database access agreement.

We therefore ask that you remove any instance where data from GISAID is used for other non-approved purposes or provided to others. As you know, GISAID data is provided to

users directly, and re-transmission or re-publication is not allowed. This mechanism ensures that the data providers themselves retain full authority over the data they share via GISAID.

As you correctly point out, GISAID supports the maintenance of phylogenetic ‘trees for many other users’ (incl. submitting labs). GISAID relies on various software programs towards tree-building, including but not limited to augur/auspice which in itself is reliant on software code developed by others, e.g., MAFFT, FastTree, IQ-TREE and TreeTime. GISAID’s support ensures those trees remain up-to-date and compliant with the terms under which data is made available. Not only does GISAID enable users to generate and download an augur json file directly from EpiCoV, GISAID’s critical support of augur/auspice was also recognized by you in 2015.

GISAID reiterates its willingness to ensure that your global subsampled phylogenetic tree remains up-to-date. Once you provide GISAID with all necessary criteria for set selection, subsampling and target size for each tree. GISAID technical staff can gladly look into the modalities of enabling the 56 trees you are making reference to.

With respect to your variant forecast dashboard using the MLR model; We note that you never requested permission to use data shared via GISAID for such a web service, nor has GISAID authorized such use. If you would like, GISAID technical staff can devise a plan to make this dashboard directly available through the GISAID platform.

We also note that the various dashboards/webservices you identified in your email imply that GISAID approved the dashboard and the data for these purposes. This is incorrect; As noted above, GISAID never provided permission for these uses. The use of the GISAID logo and statements related to GISAID on these dashboards and webservices give off the appearance that these are authorized and supported by GISAID, which is incorrect. We ask that you correct this issue immediately

Finally, we remind you that GISAID provided permission for you as an individual to use data only to maintain your global subsampled phylogenetic tree (nCoV) and generate your 'nextregion' packages for download in GISAID. Since that time, you have not been granted permission to use data from GISAID in any other form or any other dashboard and webservices.

Thank you again for your attention to this matter

GISAID Secretariat

On 10/10/25, 16:21 UTC, "Trevor Bedford"  [@bedford.io> wrote:](mailto:@bedford.io)

Dear GISAID Secretariat and dear Peter,

I'm sorry that you're not available to talk. I really do think this the more efficient use of everyone's time.

I've excerpted my response to the service support team about why the offer of packaging up a single JSON file doesn't match what we require to keep our analyses up-to-date. I'm hopeful with this technical blocker in mind, we can then continue this current email thread.

Thanks for getting in touch and I'm glad you've been able to run our Nextstrain software to provide trees for many other users. The pattern of generating and sharing Augur JSONs for viewing in [auspice.us](#) is something we also recommend to public health labs. Additionally some groups choose to share these JSONs either publicly or privately via [nextstrain.org/groups](#). We generally just hope the software is being useful. Please let me know if there are specific things that we could do to make it more useful for you.

And I hugely appreciate the specific offer of help. However, we're currently running <https://github.com/nextstrain/ncov> that produces a number of JSONs that live under nextstrain.org/ncov/gisaid/, including for example:

- <https://nextstrain.org/ncov/gisaid/global/all-time>
- <https://nextstrain.org/ncov/gisaid/north-america/all-time>
- <https://nextstrain.org/ncov/gisaid/asia/2m>
- <https://nextstrain.org/ncov/gisaid/21L/europe/6m>

This is a matrix across 7 different regions, across 4 different temporal resolutions and across 2 rootings (Wuhan and 21L). This totals 56 different JSON files that we produce weekly via the ncov workflow.

In addition we run Multinomial Logistic Regression (MLR) on clades and lineages from sequence counts as visible here: <https://nextstrain.org/sars-cov-2/forecasts>

I suspect there's also specifics in our ncov workflow that are not in your current Augur-based workflow, including things like coloring the tree from MLR fitness (https://nextstrain.org/ncov/gisaid/global/6m?c=mlr_lineage_fitness) and coloring the tree based on immune escape from DMS assay data (https://nextstrain.org/ncov/gisaid/21L/global/all-time?c=immune_escape).

For these reasons, I'd prefer flat file access to keep things updated on our end. I believe that for historical reasons you were creating a Nextstrain specific file here:

https://www.epicov.org/epi3/3p/nextflu/export/corona2020_fulldump.json.bz2

But we shouldn't need a Nextstrain specific file and should be able to use the same file as provided to other users like <https://cov-spectrum.org> and <https://cov.lanl.gov/content/index>. I also had previously written to you about sharding these files to decrease load on everyone's end (please see below email from July 2024).

What would be the process to get access to these same files that are already created for CoVSpectrum, LANL, etc...?

I repeat the primary question here. What should be the process to get access to files that are already provisioned for CoVSpectrum, LANL, etc...? If the issue was the amount of resources it takes to package up Nextstrain specific data provision, then this seems like it would resolve the issue for both of us. Am I correct?

Thank you for your time, attention and expertise.

Best,
- Trevor

On Oct 7, 2025, at 5:44 PM, GISAID Secretariat @gisaid.org> wrote:

Dear Mr. Bedford,

We are in receipt of your email related to discussions on the data packages for download.

Unfortunately, given the numerous projects undertaken by our Teams at this time, we are unable to respond with an individual discussion request. We have found that our staff resources are best allocated to address the user requests by email, as they arise within the GISAID platform.

Our Team responded to you promptly, albeit (and unfortunately) to an old email address. We have informed them of your current email, and they will respond shortly with suggestions on how GISAID can support the phylogenetic tree that you publish.

We trust that their guidance will resolve any issues that you have.
Best wishes,

GISAID Secretariat

On 10/6/25, 14:46, "Trevor Bedford" @bedford.io> wrote:

Dear GISAID Secretariat and dear Peter,

Can we please speak? There's a lot here and it will be a more efficient and more collegial use of our time if we can just speak in over the phone or over Zoom. I'm happy to send an invite.

If this works for you, please let me know some times that are preferable and I'll do my best to make this work with my schedule.

Thank you and take care,
- Trevor

On Oct 4, 2025, at 12:59 AM, GISAID Secretariat @gisaid.org> wrote:

Dear Mr. Bedford,

Thank you for your responding to our important email notice of 1-Oct-2025 regarding GISAID's decision that it will no longer make the 'nextregion' package available as it is no longer being downloaded by users in a meaningful manner.

In addition, we remind you that our email also indicated that the data provision GISAID has generated for you on a weekly basis to facilitate updates of the 'nextregion' package, will no longer be produced. This means the link '[epicov.org/epi3/...](http://epicov.org/epi3/)' that was provided to you for the past five years as an individual, i.e., a registered user, will cease to work.

Regarding your global subsampled phylogenetic tree, we propose the following mechanism by which GISAID will keep this feature up-to-date. After consulting with our staff and advisors on the feasibility of keeping your global tree up-to-date, there was a clear consensus that continuing to generate, zip and move big files back and forth is not sustainable and a waste of resources. It is recommended that you provide to GISAID the parameters for Augur and GISAID will run it for you. You would then be provided the Augur output JSON file with the relevant subsample for your phylogenetic tree.

With regard to the email from our legal counsel, we would indeed appreciate that you follow up on his questions with him directly. Accordingly, we have copied Mr. Songer on this response.

Finally, please note: We have removed your insertion into this email thread, as it not only detracts from our communication, but has no relevance to the foregoing.

Best wishes

GISAID Secretariat

From: Trevor Bedford [REDACTED]@bedford.io>
Date: Thursday, October 2, 2025 at 07:22
To: GISAID Secretariat [REDACTED]@gisaid.org>
Subject: Re: data packages made available for download

Dear GISAID Secretariat and dear Peter,

Thanks so much for getting in touch. I had completely forgotten about the outstanding email from Michael Songer from January 2023. There was a lot that I dropped in that interim. My apologies for this.

Thank you hugely for providing the backbone for ongoing genomic surveillance of SARS-CoV-2 and for keeping provisioning the flat file at www.epicov.org/epi3/3p/nextflu/export/corona2020_fulldump.json.bz2 for our continued updating of nextstrain.org/ncov/gisaid/. It's hard to believe that it's almost been six years since we started this work together in January 2020 (and still longer for work on flu).

I'd love to keep this partnership going and keep nextstrain.org/ncov/gisaid/ updating. I'd love to know if there's specific things you'd like us to be doing differently. I had noted in an email to TechDev on July 2, 2024 (copied below) about a strategy to lighten the load for provisioning data (and our processing of data), but there's many ways to do things.

But all this said, perhaps it would be best if we could have a phone call? It would be great to talk.

Separately, I can still plan to follow up on Michael Songer's queries. Should I respond directly to him or to you?

Take care,
- Trevor

>>> INSERT REMOVED <<<

On Oct 1, 2025, at 2:35 AM, GISAID Secretariat
<@gisaid.org> wrote:

Dear Mr. Bedford,

During a recent review of data packages made available for download in EpiCoV we noticed that your package 'nextregion' in the 'Genomic epidemiology' section is for the most part no longer being downloaded by users. To be precise, during the past seven months (Mar-Sep 2025) only 79 users have made a total of 88 downloads of your package 'nextregion.'

Because GISAID's resources are limited, we have made the determination to no longer make the 'nextregion' package available for download in EpiCoV.

Consequently, the data provision GISAID has generated for you on a weekly basis for the past five years to facilitate updates of the 'nextregion' package in the 'Genomic epidemiology' section, will no longer be produced.

Of course, GISAID will gladly work with you to keep up-to-date your global subsampled phylogenetic tree of the ‘Genomic epidemiology of novel coronavirus (nCoV)’ GISAID agreed to support in January 2020 (see your email update 26-Jan-2020).

Additionally, we draw your attention to an outstanding matter. We came across an email you sent on 2. January 2023 to our legal counsel, Mr. Michael Songer (see attached). You indicated at the time that you would follow up on his questions, once you are able following a medical leave. Our counsel, you will recall, had previously provided you with ‘detailed clarification on data use under the general DAA vs. web application data agreements’ you sought. We would appreciate if you provide us with your answers to our counsel’s questions, and we apologize in advance, if we have misplaced your reply.

Best wishes

GISAID Secretariat

