

Jul 09, 2020

# SARS-CoV2 GISAID submission protocol

Nabil-Fareed Alikhan<sup>1</sup>, Emma Griffiths<sup>2</sup>, Ruth Timme<sup>3</sup>, Duncan MacCannell<sup>4</sup>

<sup>1</sup>Quadram Institute Bioscience; <sup>2</sup>University of British Columbia; <sup>3</sup>US Food and Drug Administration;

<sup>4</sup>US Center for Disease Control and Prevention

**1** Works for me [dx.doi.org/10.17504/protocols.io.bh98j99w](https://dx.doi.org/10.17504/protocols.io.bh98j99w)

GenomeTrakr

Tech. support email: [genomeTrakr@fda.hhs.gov](mailto:genomeTrakr@fda.hhs.gov)



Nabil-Fareed Alikhan

Quadram Institute Bioscience

## ABSTRACT

This protocol provides the steps needed to establish a new GISAID submission environment for your laboratory. Once established, this protocol covers genome submission sample meta to GISAID.

DOI

[dx.doi.org/10.17504/protocols.io.bh98j99w](https://dx.doi.org/10.17504/protocols.io.bh98j99w)

## PROTOCOL CITATION

Nabil-Fareed Alikhan, Emma Griffiths, Ruth Timme, Duncan MacCannell 2020. SARS-CoV2 GISAID submission protocol. **protocols.io**

[dx.doi.org/10.17504/protocols.io.bh98j99w](https://dx.doi.org/10.17504/protocols.io.bh98j99w)

## KEYWORDS

metadata, INSDC, ERC000033, ENA, EBI, SARS-Cov2, COVID-19

## LICENSE

————— This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

## CREATED

Jul 07, 2020

## LAST MODIFIED

Jul 09, 2020

## PROTOCOL INTEGER ID

38944

## GISAID Submissions

- 1 GISAID submissions require contextual information (metadata) and the COVID-19 consensus sequence or genome sequence. It is much simpler than other databases, as it has no hierarchical structure. In GISAID, one set of contextual information links to one genome assembly.

We advise that contextual information should be in line with the **PHA4GE metadata specification**:

<https://github.com/pha4ge/SARS-CoV-2-Contextual-Data-Specification>

GISAID has its own extensive documentation, which is available once you have registered. You are encouraged to write to the GISAID curators if there are any issues. This protocol, therefore, is a brief overview to help you get started.

- 1.1 You can register on GISAID here: <https://www.gisaid.org/registration/register/>

There are strict terms on your rights to your data, and how you can use other submitted data. Please read these terms carefully. There will be a registration form to fill

folder, too; the confirmation email will have your full name in the subject line.

Personal Data	
Mr/Ms*	<input type="radio"/> Ms <input type="radio"/> Mr
Salutation	<input type="text"/>
First name*	<input type="text"/>
Middle name	<input type="text"/>
Last name*	<input type="text"/>
Title	<input type="text"/>
Desired Login ID*	<input type="text"/>
Contact Information	
Use your company or research institute's email! Web emails such as Gmail or Yahoo will lead to significant delays in processing.	
Institution*	<input type="text"/>
Department	<input type="text"/>
Street 1*	<input type="text"/>
Postal code	<input type="text"/>
City*	<input type="text"/>
Location*	<input type="text" value="Select location"/>
State/province	<input type="text"/>
Telephone*	<input type="text"/>
Fax	<input type="text"/>
Mobile	<input type="text"/>
E-Mail*	<input type="text"/>

#### Data Access Agreement

GISAID Registration form

Once submitted, your registration will be moderated and you should receive an email notification. Your registration request is manually inspected, so this may take some time.

## 1.2 GISAID Submission methods

There are two submission methods;

1. Uploading single samples.
2. Batch upload of multiple samples.

In both cases, we advise that contextual information should be in line with the **PHA4GE metadata specification**.

A brief overview of the fields and guidance are tabulated below. Extended description of the specification is available here: <https://github.com/pha4ge/SARS-CoV-2-Contextual-Data-Specification>

These fields below are drawn from the GISAID submission form, with description and mapping to the respective PHA4GE field should help you transfer your metadata from the PHA4GE table to the something acceptable for GISAID submission.

#### Description of GISAID fields

GISAID Batch Submission Headers (as of 2020-06-19)	GISAID Definition	GISAID Requirement Status
Submitter	enter your GISAID-Username	mandatory
FASTA filename	the filename that contains the sequence without path (e.g. all_sequences.fasta not c:\users\meier\docs\all_sequences.fasta)	mandatory

Virus name	e.g. hCoV-19/Netherlands/Gelderland-01/2020 (Must be FASTA-Header from the FASTA file all_sequences.fasta)	mandatory
Type	default must remain "betacoronavirus"	mandatory
Passage details/history	e.g. Original, Vero	mandatory
Collection date	Date in the format YYYY or YYYY-MM or YYYY-MM-DD	mandatory
Location	e.g. Europe / Germany / Bavaria / Munich	mandatory
Additional location information	e.g. Cruise Ship, Convention, Live animal market	
Host	e.g. Human, Environment, Canine, Manis javanica, Rhinolophus affinis, etc	mandatory
Additional host information	e.g. Patient infected while traveling in ....	
Gender	Male, Female, or unknown	mandatory
Patient age	e.g. 65 or 7 months, or unknown	mandatory
Patient status	e.g. Hospitalized, Released, Live, Deceased, or unknown	mandatory
Specimen source	e.g. Sputum, Alveolar lavage fluid, Oro-pharyngeal swab, Blood, Tracheal swab, Urine, Stool, Cloakal swab, Organ, Feces, Other	
Outbreak	Date, Location e.g. type of gathering, Family cluster, etc.	
Last vaccinated	provide details if applicable	
Treatment	Include drug name, dosage	
Sequencing technology	e.g. Illumina Miseq, Sanger, Nanopore MinION, Ion Torrent, etc.	mandatory
Assembly method	e.g. CLC Genomics Workbench 12, Geneious 10.2.4, SPAdes/MEGAHIT v1.2.9, UGENE v. 33, etc.	
Coverage	e.g. 70x, 1,000x, 10,000x (average)	
Originating lab	Where the clinical specimen or virus isolate was first obtained	mandatory
Address		mandatory
Sample ID given by the sample provider		
Submitting lab	Where sequence data have been generated and submitted to GISAID	mandatory
Address		mandatory
Sample ID given by the submitting laboratory		
Authors	a comma separated list of Authors with complete First followed by Last Name	

Description of GISAID fields

#### PHA4GE guidance on GISAID submission fields.

GISAID Batch Submission Headers (as of 2020-06-19)	PHA4GE Field	PHA4GE Guidance
Submitter	*user inputs	The submitter must acquire an account and provide their username in this field.
FASTA filename	fasta filename	This field can be populated by the PHA4GE field "fasta filename".

Virus name	isolate	While the meanings and structures of the GISAID field "Virus name" and the PHA4GE field "isolate" overlap, GISAID requires a slightly different structure for Virus name. e.g. PHA4GE structure: SARS-CoV-2/country/sampleID/2020, GISAID structure: hCov-19/country/sampleID/2020. Change "SARS-CoV-2" to "hCov-19" in the isolate name.
Type	*should always be betacoronavirus	Provide "betacoronaviruses".
Passage details/history	specimen processing; lab host; passage number; passage method	This field can be populated by the PHA4GE fields "specimen processing", "lab host", "passage number" and "passage method". If the information is unknown or can not be shared, put "unknown".

Collection date	sample collection date	This field can be populated by the PHA4GE field "sample collection date". Caution: collection date may be considered public health identifiable information. If this date is considered identifiable, it is acceptable to add "jitter" to the collection date by adding or subtracting a calendar day (acceptable by GISAID). Do not change the collection date in your original records. Alternatively, "received date" may be used as a substitute in the data you share. The date should be provided in ISO 8601 standard format "YYYY-MM-DD".
Location	geo_loc_name (country)	This field can be populated by the PHA4GE field "geo_loc_name (country)".

Additional location information	exposure event	This field can be populated by the PHA4GE field "exposure event". Caution: this may be sensitive information. Consult the data steward before sharing. If the information is unknown or can not be shared, leave blank or put "unknown".
Host	host (scientific name)	This field can be populated by the PHA4GE field "host (scientific name)".
Additional host information	N/A	If the information is unknown or can not be shared, leave blank.
Gender	host gender	This field can be populated by the PHA4GE field "host gender". Caution: the host gender may be considered public health identifiable information. Consult the data steward before sharing. If the information is unknown or can not be shared, put "unknown".

Patient age	host age	This field can be populated by the PHA4GE field "host age". Provide age in years. Age-binning is not accepted. Caution: the host age may be considered public health identifiable information. Consult the data steward before sharing. If the information is unknown or can not be shared, put "unknown".
Patient status	host health state	This field can be populated by PHA4GE field "host health state". If the information is unknown, or can not be shared, put "unknown".

Specimen source	anatomical material; anatomical part; body product; environmental material; environmental site; collection device; collection method	This field can be populated by the PHA4GE fields "anatomical material", "anatomical part", "body product", "environmental material", "environmental site", "collection device" and "collection method". Separate the values using semi-colons. If the information is unknown or can not be shared, leave blank or put "unknown".
Outbreak	N/A	If the information is unknown or can not be shared, leave blank.
Last vaccinated	N/A	If the information is unknown or can not be shared, leave blank.
Treatment	N/A	If the information is unknown or can not be shared, leave blank.
Sequencing technology	sequencing instrument	This field can be populated by the PHA4GE field "sequencing instrument". If the information is unknown or can not be shared, put "unknown".



Assembly method	consensus sequence method	This field can be populated by the PHA4GE field "consensus sequence method". If the information is unknown or can not be shared, leave blank or put "unknown".
Coverage	assembly coverage depth	This field can be populated by the PHA4GE field "assembly coverage depth". If the information is unknown or can not be shared, leave blank or put "unknown".
Originating lab	sample collected by	This field can be populated by the PHA4GE field "sample collected by". Caution: if the name of the lab reveals geographic information, this may be considered public health identifiable information. Consult the data steward before sharing. If the information is unknown or can not be shared, leave blank or put "unknown".

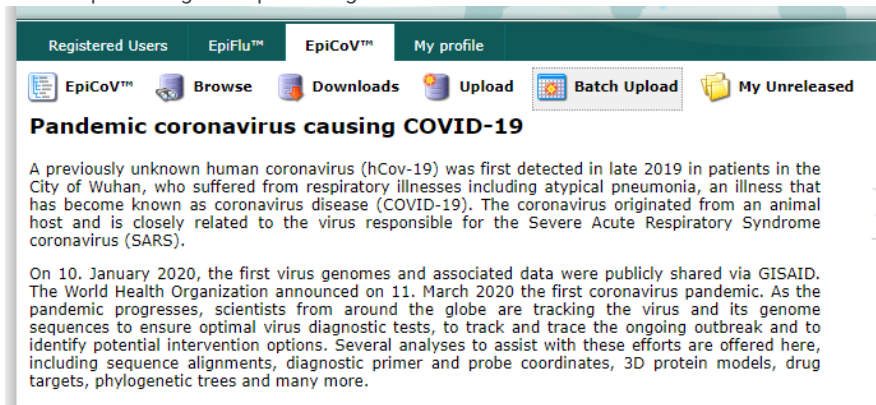
Address	sample collector contact address	This field can be populated by the PHA4GE field "sample collector contact address".
Sample ID given by the sample provider	specimen collector sample ID	This field can be populated by the PHA4GE field "specimen collector sample ID". Caution: the sample ID may be considered public health identifiable information. Consult your data steward before sharing. You may need to provide an alternative ID.
Submitting lab	sequence submitted by	This field can be populated by the PHA4GE field "sequence submitted by". If information is unknown or can not be shared, leave blank or put "unknown".
Address	sequence submitter contact address	This field can be populated by the PHA4GE field "sequence submitter contact address". If information is unknown or can not be shared, leave blank or put "unknown".

Sample ID given by the submitting laboratory	consensus sequence ID	This field can be populated by the PHA4GE field "consensus sequence ID". If information is unknown or can not be shared, leave blank or put "unknown".
Authors	authors	This field can be populated by the PHA4GE required field "authors". If information is unknown or can not be shared, leave blank or put "unknown".

PHA4GE guidance on GISAID submission fields.

## 2 Uploading single samples

You can upload single sample through a web form shown below.



GISAID dashboard once you've logged in.

**Single Upload**

Enter and upload genetic sequence and metadata, available clinical and epidemiological data, geographical as well as species-specific data. Data will be reviewed by a curator prior to release. An email confirmation will be issued upon release.

**Virus detail**

Virus name\*

Accession ID

Type

Passage details/history\*

**Sample information**

Collection date\*

Location\*

Additional location information

Host\*

Additional host information

Gender\*

Patient age\*

Patient status\*

Specimen source

Outbreak Detail

Last vaccinated

Treatment

Submission form for single samples

### 3 Uploading multiple samples (Batch upload)

You must first explicitly request the batch upload feature from the curators. This will give you access to the batch upload page. Here you can upload multiple samples (metadata and sequences). There is a link to detailed documentation and templates at the bottom left.

EpiCoV™ Browse Downloads Upload Batch Upload My Unreleased

**GISAIID hCoV-19 Batch Upload**

Upload genetic sequence as single FASTA-File and metadata, available clinical and epidemiological data, geographical as well as species-specific data as XLS. Data will be reviewed by a curator prior to release. An email confirmation will be issued upon release.

Metadata as Excel\*

max size: 5M  No file chosen

Sequences as FASTA\*

max size: 32M  No file chosen

Report

Batch upload page