Questionnaire for databases

Background

1. Name of the database

GISAID EpiFlu™ Database

2. When was your database established?

The launch of GISAID's EpiFlu™ database on the occasion of the 61st World Health Assembly (2008) marked the culmination of the establishment of the Global Initiative on Sharing All Influenza Data, a collaboration involving representatives of Member States, scientists of the Global Influenza Surveillance and Response System (GISRS) and Global Influenza Programme (GIP) of the WHO, and experts in licensing intellectual property.[1]

3. Who hosts and manages the database?

The Federal Republic of Germany is the official host of the GISAID platform and database in a public-private partnership with the administrative arm of GISAID 'Freunde von GISAID e.V., a registered non-profit association which ensures oversight of data access compliance matters to protect the interests of both data providers and the user community.

Since 2010, the German government provides the technical hosting facilities through its Federal Office for Food & Agriculture (BLE) and validation of Data through its Federal Research Institute for Animal Health, the Friedrich-Loeffler-Institute (FLI).

Scientific oversight of GISAID is provided by directors of all six WHO Collaborating Centres for Reference and Research on Influenza (CCs), directors of WHO National Influenza Centres (NICs) and directors of leading FAO/OIE Reference Laboratories for Avian Influenza, through GISAID's Scientific Advisory Council (SAC). The SAC also promotes GISAID's scientific mission, fostering scientific collaboration and capacity building in developing countries. The WHO Global Influenza Programme has observer status in the GISAID SAC.

Technical oversight is provided by the Database Technical Group (DTG), with expertise in genetic sequencing and data analysis. The DTG guides GISAID's database development and its members assist with training during WHO-sponsored training visits to CCs and in GISAID-WHO bioinformatics workshops on genetic analyses of human and animal influenza viruses. This educational programme is essential to enable countries to optimise their benefit from the Data on their viruses.

---

[1] Elbe, S., and Buckland-Merrett, G. (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*, 1: 33–46. doi: 10.1002/gch2.1018

Declaration of Interest (DOI Statement):

GISAID declares that at no time since its formation in 2008 has GISAID or its management, or any of its board members, received any research support from, investment interests of any kind in, or performed any form of contract work for, industry or any commercial entity.[2]

The public-private-partnership of the GISAID Initiative is supported by financial and in-kind contributions from WHO Member States and the private sector. Additional support in the form of grants and technical partnerships strengthen the Initiative's important role in global health. Both operational and scientific independence remain unaffected.

GISAID does not own intellectual property that might be enhanced or diminished by the outcome of the implementation of Decision WHA70(10)8(b): *Analysis of the implications of pursuing or not pursuing possible approaches to genetic sequence data under the PIP Framework*, requested by Member States at the 70th World Health Assembly, or work of the PIP Advisory Group.

---

[2] "Commercial entity" includes any commercial business e.g. pharmaceutical industry, an industry association, research institution or other enterprise of which funding is significantly derived from commercial sources with an interest related to the subject of the meeting or work.

1. Organization and functioning of the database

4. Please describe the sequencing data (e.g. source material, type, etc.) stored in your databases.

Genetic Sequence Data in GISAID is derived from influenza viruses (types A, B and C) in many types of source material, including e.g. original specimens from human and animal (avian and mammalian) hosts (including swabs and environmental samples), or influenza viruses isolated and passaged in cell culture, or in embryonated hen eggs, and candidate vaccine viruses (CVV). Sequence data is categorized and stored by virus isolate ID (EPI_ISL_XXXXX). A unique virus isolate is defined by a combination of Virus Name, Passage History, Date of Harvest and Submitting Laboratory. Sequence data for each segment is assigned a unique accession number (EPIXXXX).

5. Does your database also store/provide access to associated Data (e.g. epidemiological data)? If so, please provide a description.

Yes. In addition to Genetic Sequence Data, GISAID stores and provides >30 fields of associated metadata, both epidemiological and clinical, most of them searchable (s)

1. Virus Name
2. Virus Type (s)
3. Virus Subtype (s)
4. Lineage of Influenza B (s)
5. Date of Specimen Collection (s)
6. Specimen Source
7. Host Information (e.g. Animal Species) (s)
8. Location of Collection (s)
9. Passage History of Virus Isolate (s)
10. Name and Contact Details of Originating Laboratory (s)
11. Originating Sample ID
12. Name and Contact Details of Submitting Laboratory (s)
13. Submitting Sample ID
14. Date of Data Submission (s)
15. Name of (with ability to contact) Individual Submitter of the Data;
16. Antiviral Susceptibility (s)
17. Antigenic Characterization
18. Author
19. PMID
20. INSDC Accession Numbers (in case of imported data)

For human samples/isolates:

21. Patient Age
22. Patient Gender
23. Patient Health Status
24. Previous Vaccination History
25. Outbreak Information
26. Antiviral Treatment

For animal samples/isolates:

27. Host Subgroup/Family/Species
28. Domestic Status

29. Health Status
30. In-Vivo Pathogenicity Test (avian)
31. Vaccination Status
32. Strain used for Vaccination

6. How many influenza genetic sequences are stored in your database?

GISAID's EpiFlu™ Database contains about one million nucleotide sequences of genome segments of approximately 250,000 influenza viruses, approximately three-quarters of which are human viruses.

Data from over 1,000 institutions are entrusted to and protected by GISAID's sharing mechanism, which is governed by the GISAID EpiFlu™ Database Access Agreement.

7. How many sequences from influenza viruses with human pandemic potential[3] are stored in your database?

GISAID's EpiFlu™ Database stores and makes publicly available:

Sequences of approximately 17,000 genome segments of over 2,500 Influenza Viruses (isolates/cases) with Human Pandemic Potential (IVPP); three-quarters of the Data for IVPP identified since 2013 were submitted directly to EpiFlu™.

Human IVPP viruses of different subtypes:
- 682 H5N1 viruses (3,531 sequences)
- 23 H5N6 viruses (176 sequences)
- 1 H6N1 virus (8 sequences)
- 1 H7N1 virus (1 sequence)
- 3 H7N2 viruses (17 sequences)
- 2 H7N3 viruses (16 sequences)
- 58 H7N7 viruses (190 sequences)
- 1,444 H7N9 viruses (10,744 sequences)
- 30 H9N2 viruses (167 sequences)
- 5 H10N8 viruses (40 sequences)
- 35 H1N1v viruses (336 sequences)
- 11 H1N2v (108 sequences)
- 251 H3N2v viruses (1,606 sequences)

In addition to the above there are laboratory-derived candidate vaccine viruses (CVV) derived from animal or environmental viruses of relevant subtypes shared with GISRS in the context of the PIP Framework: 19 H5N1; 2 H5N6; 1 H5N8; 2 H7N1; 1 H7N2; 1 H7N3; 1 H7N7; 1 H9N2.

---

[3] The PIP Framework defines 'influenza viruses with human pandemic potential' as any wild-type influenza virus that has been found to infect humans and that has a haemagglutinin antigen that is distinct from those in seasonal influenza viruses so as to indicate that the virus has potential to be associated with pandemic spread within human populations with reference to the International Health Regulations (2005) for defining characteristics. (see PIP Framework Section 4.2)

Total numbers of animal and environmental viruses of IVPP subtypes for which sequences are available: 7,226 H5N1; 1369 H5N6; 959 H5N8; 529 H6N1; 396 H7N1; 501 H7N2; 921 H7N3; 1,207 H7N7; 1,117 H7N9; 5,553 H9N2; 118 H10N8.

8. What information/annotation is provided about the sequences?

Information	**Yes**	**No**
Submitting Laboratory/Originating laboratory	Yes	
Source material	Yes	
Country of Origin of the Data	Yes and linked to Submitting Laboratory	
Country of origin of the source material	Yes	
Date of submission	Yes	

Other:
In addition to the meta data which is at the isolate level, each gene segment sequence includes:

1. Segment Identifier (entered by submitter)
2. RNA Segment Designation (entered by submitter then checked by GISAID annotation system)

Each sequence is assigned the following annotations by the GISAID annotation system:

3. Segment Accession Number;
4. Influenza Type (entered by submitter then checked by GISAID annotation system);
5. Subtype and Lineage (entered by submitter then checked by GISAID annotation system);
6. Protein Sequences (open reading frames);
7. Length of nucleotide and protein coding sequences;
8. Completeness of coding sequence

9. Briefly explain the process to upload Data in your database.

All sequence Data are passed through a distinct annotation and curation process during upload, to ensure validity and the best possible Data quality.

There are currently two distinct ways to upload Data to GISAID.

1. Single Upload is possible via a web-based step-by-step protocol with description of the entry fields and procedures using free text, drop down and categorical fields.

2. The Upload of multiple datasets is realized via a batch upload facility using a Microsoft Excel® template (provided by GISAID) with description of fields, followed by a web-based upload procedure giving feedback of potential mistakes or incomplete Data and providing automatic annotation of sequences (virus type and subtype, segment designation, protein coding sequences during upload.

Release and editing of entries is done autonomously by the submitter. All submitters and users are able to seek assistance from the qualified GISAID EpiFlu™ Database Service Team, free of charge.

10. Briefly explain how Data is accessed by users and the general public.

Every natural person (without exception) wishing to access Data in GISAID engages in a one-time registration process, whereby the person's complete name, affiliation and contact details (address, telephone, email) are provided, and GISAID's Terms of Use are agreed.

Following an automated and manual review to achieve positive verification of the identity and validity of the registration information provided, individual Users receive their personal and unique User Access Credentials (username and password) to provide access to all Data contained within GISAID following a standard login procedure.

The use of personal User Access Credentials for each and every individual permits the system to associate Data with individuals both submitting or accessing Data in GISAID.

Additionally, Users are provided with integrated search functions that allow easy retrieval of the desired datasets. Data can be downloaded in several formats and analyzed with either basic or more sophisticated tools. The database provides direct access to analysis tools for more detailed general and influenza-specific analyses.

11. From which countries do most sequences in your database originate?

Proportions of virus sequence data in GISAID originating from countries in different continents:

34%	Asia
31%	North America
23%	Europe
5%	Oceania
4%	Africa
4%	South America

12. To the best of your knowledge, on average, how quickly after sequencing are sequences uploaded to your database?

The majority of Data submitted to GISAID is made publicly available prior to the publication of a scientific study. Thus, in general GISAID's EpiFlu™ Database receives submissions of Data from current/novel strains significantly faster than other influenza data repositories. As the volume of Data has increased, the delay between specimen collection and submission of sequence Data directly to GISAID has decreased. Data on seasonal human influenza viruses for the biannual WHO vaccine strain consultation meetings (VCM) are deposited by WHO CCs and NICs within a time-frame of days to a few weeks of sequencing, depending on urgency and other circumstances.

For example, during 2018 direct submissions to GISAID have accounted for more than 94% of virus Data submitted within 3 months of specimen collection. During 2016, for Europe and N. America about 60% of complete HA sequences were submitted to EpiFlu™ within 2 months of specimen collection.[4]

During outbreaks of novel zoonotic infections uploading can be very timely, e.g. in the case of A/H7N9, in March 2013, Data were made publicly available through GISAID in less than 48 hours after sequencing. Of genetic Data made available during 2016-17 for H5N6, H5N8 and H7N9 subtypes from all hosts, 98% of Data submitted within 30 days of specimen collection were unique to EpiFlu™. Such 'near-real time' deposits are essential for risk assessment of, and effective response to, zoonotic infection.

---

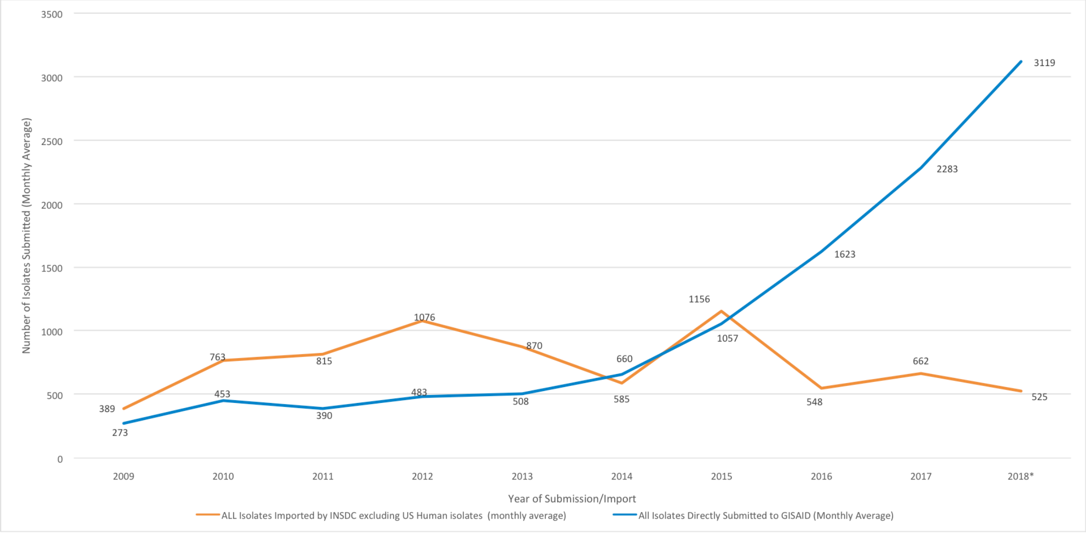
[4] Morris et al., *Trends Microbiol*. **26**, 102-118, 2018

13. How many sequences are uploaded into and downloaded from your database on a monthly basis?

Over the full period of EpiFlu™ operation (2008 to 2018) the average upload of sequences on a monthly basis is of the order of 8,600 per month.

The Figure illustrates a continuing annual increase of >40% in the average monthly upload of virus data submitted directly to GISAID (blue graph), from approximately 500 'viruses' per month in 2013 to over 3,000 'viruses' per month in 2018, while the influenza Data contributed from INSDC declined (orange graph).

From 1 January to 30 April 2018, the corresponding monthly average number of sequences (of individual segments) uploaded directly to GISAID was over 14,000/month.



14. Who are the principal users of your database?

Users	**Yes**	**No**
General Public		No
Academic Institutions	Yes	
Researchers	Yes	
Industry	Yes	
Unidentified User		Not permitted

GISAID has 8,000 active participants and well over 1,000 institutions contributing Data

Policy

15. Does your database have an access policy?

Data Access policy	**Yes**	**No**
	Yes	
While access to Data in GISAID is free-of-charge, it is accessible only to those who identify themselves and agree to GISAID's basic premise of upholding a scientific etiquette, e.g. acknowledging the Originating Laboratories providing the specimen and Submitting Laboratories who generate the sequence Data, ensuring fair exploitation of results from analyses of the Data, and attaching no restrictions to Data that have been submitted to GISAID. Access to the GISAID Database is governed by a legally-binding Database Access Agreement to which every User must agree. A User must be a natural person. All bona fide Users that have successfully identified themselves and accepted the GISAID Terms of Use are provided personal access credentials to the database. Access credentials are subject to revocation should a violation of Terms of Use occur. Prior to the issuance of access credentials, the identity of each and every user is confirmed by a team of analysts that adhere to a SOP, to ensure that access to and use of Data in GISAID are compliant with the license terms of the Database Access Agreement, setting forth specific conditions on the use of Data, i.e. rights and responsibilities. While the GISAID Database is publicly accessible and designated by editorial boards as providing Open Access, the Data in GISAID does not fall under the legal definition of Public-Domain where preexisting intellectual property rights to the Data are relinquished. GISAID was launched to provide an alternative to Public-Domain databases (e.g. INSDC), where intellectual property rights are relinquished and access takes place anonymously with no verifiable means to enforce the rights of Data contributors.		

16. If so, does your Data access policy cover the following:

Data Access and Use	**Yes**	**No**
Access to the database (registration, identification)	see #15	
Use of Data for scientific purposes	as follows	
Data in GISAID is licensed for the conduct of analyses of the Data and the publication of such analyses (not the Data), provided the publication acknowledges the contributors i.e. the Originating Laboratory where the clinical specimen or virus isolate was first obtained and the Submitting Laboratory where the sequence data used in the analysis have been generated and submitted to the EpiFlu™ Database. Failure to comply could result in sanctions such a restricted access to the GISAID database.		
Use of Data for commercial purposes	as follows	
Data in GISAID is licensed for the development, testing and dissemination of interventions such as vaccines, diagnostics and therapeutics. Use beyond these purposes is not licensed through GISAID. For example, Data in GISAID may not be subjected to any changes of ownership, such as the placement of any other rights onto the Data in whole or fraction thereof. (see IP Rights or other restrictions)		
Further sharing of downloaded Data with a third-party		Not Permitted

Uploading of downloaded Data to another database		Not Permitted
Acknowledgment of originating laboratory/country	Required	
Collaboration with originating laboratory/country	Best Efforts Required	
Intellectual property rights or other restrictions on the Data		Not Permitted

Any alteration of IP Rights to Data in GISAID is not permitted.

When Data are deposited in the GISAID database, all inherent rights, e.g. IP Rights attached to the Data, remain in place. These IP Rights are explicitly preserved and may not be altered under the license provided through GISAID's Terms of Use.

GISAID Users have agreed not to offer, impose or attach any terms on the Data that alter the ownership and any rights to the Data. Subject only to any pre-existing third party rights on the Data, Users have acknowledged and agreed that all Data will be freely shared among and used by all other Authorized GISAID Users.

It is the User's sole responsibility to obtain any additional authorization or license from the owners of the Data, should it be necessary for Use of the Data, which has not been addressed by the License offered through GISAID.

Suspension/Termination of access to the database	Yes	

Violation of GISAID's Terms of Use will trigger temporary suspension of access, or possibly revocation of access credentials to the GISAID EpiFlu™ database, to protect the integrity of the sharing mechanism and its user base.

Other/Comments:

GISAID Data has become the most frequently used source of influenza data in peer-reviewed publications.

The essential partnership between GISAID and the Global Influenza Surveillance and Response System (GISRS) and Global Influenza Programme (GIP) of the WHO was acknowledged during the 65th anniversary celebrations of GISRS in 2017.

In 2017, the Health Ministers of the G20 recognized the importance of GISAID in regard to virus data sharing in their Berlin Declaration.

The peer reviewed fact-finding and scoping study, commissioned by the Secretariat of the Convention on Biological Diversity (CBD), to clarify digital sequence information on genetic resource objectives of the CBD and the Nagoya Protocol, also highlights the exemplary role of GISAID in Access & Benefit Sharing: '*GISAID provides useful lessons and insights to ABS discussions' (CBD Secretariat Fact-Finding and Scoping Study on Digital Sequence Information 2018).* Sharing data through GISAID meets >75% of criteria for non-monetary benefits, and several monetary benefits listed in the Nagoyal Protocol Annex *(Text of the Nagoya Protocol Annex. Monetary and Non-monetary Benefits)*

Another example of the way in which the GISAID sharing mechanism encourages benefit sharing is provided by the Collaboration on H5 Antigenic Cartography whereby the providers of the H5N1 data will receive products of the research, synthetic HA constructs, reverse genetics viruses, and antisera, from the H5 consortium. *(Collaboration on H5 Antigenic Cartography H5 Consortium (2017))*

In parallel with the increased use of GISAID's sharing mechanism, and to enhance capacity building in developing countries, GISAID continues to collaborate with the WHO Global Influenza Programme and the Global Influenza Surveillance and Response System (GISRS), isirv (International Society for Influenza and other Respiratory Virus Diseases) and a number of institutions around the world to organise bioinformatics training workshops on the use of EpiFlu™ for genetic analyses of human and animal influenza viruses.

On its 10th Anniversary in May 2018, GISAID received many accolades for its achievements in promoting the sharing of influenza virus genetic sequence Data, a selection of which can be found here.