Pandemic Influenza Preparedness Framework

Genetic Sequence Data

WHO Survey on the Sharing of Genetic Sequence Data of Influenza Viruses with Human Pandemic Potential

Results and Analysis of data received

27 April 2016

World Health Organization

Table of Contents

Acknowledgement 3
Acronyms 4
Executive Summary 5
Background 8
Methodology Summary 9
IVPP GSD Sharing in Numbers (as of October 2014) 10
Survey Results 11
1. Mechanisms for sharing of IVPP GSD 11
2. Ease of sharing 13
3. Systematic sharing 13
4. Rapid and timely sharing of IVPP GSD 15
5. Completeness of data 18
6. Traceability 19
7. Acknowledgement and collaboration 20
8. Quality control 20
Annex 1 22
Annex 2 24
Annex 3 30

Acknowledgement

WHO would like to acknowledge the data providers and data users who completed the questionnaire and thank them for their invaluable input.

Acronyms

GISAID: Global Initiative on Sharing All Influenza Data

GISRS: Global Influenza Surveillance and Response System

GSD: Genetic Sequence Data

INSDC: International Nucleotide Sequence Database Collaboration

IRD: Influenza Research Database

IVPP: Influenza viruses with human pandemic potential

IVTM: Influenza Virus Traceability Mechanism

PIP: Pandemic Influenza Preparedness

PIP AG: PIP Advisory Group

WHO: World Health Organization

Executive Summary

Survey

As part of its guidance to the Director-General under PIP Framework Section 5.2.4, the PIP Advisory Group recommended in October 2014 a process to identify the optimal characteristics of a system for the sharing of influenza viruses with human pandemic potential under the Framework.[1] In order to inform this process, the Secretariat developed a questionnaire that was shared with laboratories that sequence IVPPs and laboratories that use the GSD to conduct research and public health risk assessments or to develop influenza vaccines and other products.

Mechanisms to share IVPP GSD

The survey confirms that data providers and users rely on 4 different databases to share and access IVPP GSD: GISAID EpiFlu™, INSDC, IRD and OpenFluDB. A majority of data providers use GISAID EpiFlu™ and several providers use more than one platform to share IVPP GSD. Data providers cited two main reasons for choosing GISAID EpiFlu™: convenience of data upload and user interface, and data access and usage policy. Providers who upload to INSDC, IRD or OpenFluDB indicated choosing these databases mainly because they are perceived as offering the broadest access possible to the research and public health community.

A majority of users access IVPP GSD through more than one database. Individual comments from users of IVPP GSD indicate that they favour different databases depending on the nature of their work. For example, laboratories that promptly need new sequences for risk assessment during outbreaks praised GISAID EpiFlu™, in particular in relation to H7N9 sequences. On the other hand, several respondents have pointed out OpenFluDB's flexibility for sequence data parsing, or IRD's tools and their value for users with limited bioinformatics capabilities. Taken together, these observations suggest that the various databases each fulfil different needs of users, whether it is rapid availability of data for public health purposes, bioinformatics features for academic research projects, or convenience.

A significant number of data providers indicated that they do not upload all IVPP GSD to publicly-accessible databases. Several GISRS laboratories mentioned that they will not upload IVPP GSD unless they receive permission from the sample-submitting laboratory. This is consistent with the core terms of reference for WHO Collaborating Centres[2], which state:

> "WHO Collaborating Centres for influenza 'upload available haemagglutinin, neuraminidase and other gene sequences of A(H5) and other influenza viruses with pandemic potential to a publicly-accessible database in a timely manner but no later than three months after sequencing is completed, **unless otherwise instructed by the laboratory or country providing the clinical specimens and/or viruses (Guiding Principle 9)**'." [emphasis added]

[1] See PIP Framework Advisory Group, Report to the Director-General, available at http://www.who.int/influenza/pip/pip_ag_oct2014_meetingreport_final_7nov2014.pdf?ua=1.
[2] See PIP Framework, Annex 5, Core term of reference 5, available at http://apps.who.int/iris/bitstream/10665/44796/1/9789241503082_eng.pdf

We do not know from the answers provided to this survey how many sequences this pertains to. Most providers, however, indicated that they share all IVPP GSD that they generate. If IVPP GSD is not uploaded to a publicly-accessible database, it is shared back with the sample-submitting laboratory, usually by email or through reports. Again, this is consistent with the PIP Framework expectation that "genetic sequence data, and analyses arising from that data, relating to H5N1 and other influenza viruses with human pandemic potential should be shared in a rapid, timely and systematic manner **with the originating laboratory and among WHO GISRS laboratories**". [emphasis added] Therefore, some IVPP GSD are shared only with the originating laboratory, GISRS, and/or WHO via emails, reports or other means; they are not uploaded to any database.

Timely access to IVPP GSD

Several data users reported being unsatisfied with timeliness of access to new IVPP GSD. Many of these laboratories are involved in risk assessments during outbreaks, in which case timely access may be more important than in other contexts. Many users, however, explained that this was only true for a limited number of viruses and that the majority of IVPP GSD are available when needed. Data providers indicated that pending publications and approval from originating laboratories or ministries of health were the main reasons for delays in sharing.

One interesting finding from the survey is the lack of mechanisms to notify users when new IVPP GSD are available. Many users simply perform manual searches at regular intervals or when outbreak events or new cases are reported. This might also have a significant impact on the timely access to IVPP GSD.

Completeness and quality of IVPP GSD

Almost all users reported being satisfied with the general quality of the data available. However, several users stated that completeness of metadata could be improved, especially for clinical data. Since both data providers and users indicated that databases were collecting this additional information, the issue might be at the provider level.

Traceability

In order to support the sharing of benefits arising from the use of IVPP GSD, identifying entities that use IVPP GSD to develop vaccines and other pandemic products is essential. To achieve this, the origin of the data must be identified properly.

When publishing results, almost all users indicated they identify the origin of the data using database accession numbers, which is very useful in order to trace use of IVPP GSD. However, results from the survey indicate that data users cannot always find information on the origin of the data (e.g. the laboratory that sequenced or detected the virus). If IVPP GSD cannot be traced back to the laboratory that sequenced or detected the virus, acknowledgement and collaboration may be made more difficult.

Acknowledgement and collaboration

The majority of providers reported that their contributions were not properly acknowledged in publications that used IVPP GSD generated by their laboratories. However, no provider reported that this made them more reluctant to share IVPP GSD.

In their comments, several data users described their current practices for acknowledging data providers. If the research focuses on a small number of new sequences, data users directly acknowledge the contributing laboratory. If, however, the research is based on well-established reference sequences or a very large number of sequences, only the accession numbers are mentioned. For many users, listing accession numbers was considered appropriate acknowledgment.

Although a majority of data users indicated that they collaborate with the data provider, many users mentioned that it was sometimes difficult to develop a project or to find funding for such collaborations.

Background

The PIP Framework is an international arrangement adopted in 2011 by the 194 Member States of WHO, and aims to improve global pandemic influenza preparedness and response, by:

1. improving and strengthening the sharing of IVPP[3] through a WHO-coordinated network of influenza laboratories (known as 'GISRS'), and;
2. promoting the fair and equitable access, by developing countries, to the benefits arising from such sharing.

During PIP Framework negotiations, Member States recognized the importance of genetic sequence data for pandemic preparedness and response and requested that the Director-General seek advice from the PIP AG on the "best process for further discussion and resolution of issues relating to the handling of genetic sequence data from H5N1 and other influenza viruses with human pandemic potential as part of the Pandemic Influenza Preparedness Framework."

In its October 2014 meeting report to the Director-General[4], the PIP AG made the following observations:

> "a. Laboratories should continue to share the GSD of influenza viruses with pandemic potential (IVPP) as soon as it becomes available because it is necessary for timely and comprehensive pandemic risk assessment and response.
>
> b. In accordance with Section 6.3.2, laboratories using GSD will meet appropriate biosafety guidelines (WHO Laboratory Biosafety Manual, 3rd edition) and employ laboratory protection best practices.
>
> c. The objective of benefit-sharing may be met by mechanisms related to monitoring products generated using influenza GSD, rather than by monitoring use of GSD and/or tracing GSD, noting that source identification is critical.
>
> d. Closer collaboration regarding open sharing of influenza GSD among the many different databases is desirable."

Thus, in its guidance to the Director-General, the PIP AG recommended a process to identify "the optimal characteristics of a system for the handling of IVPP GSD under the Framework, including consideration of data sharing systems that are best suited to meet the objectives of the Framework considering obligations and timeliness of data submission, quality assurance of data, completeness of data annotation, ease of access to data, sustainability and security of the system".

In order to inform this process, the Secretariat conducted a survey of current systems in place for the sharing of influenza genetic sequence data.[5] This report contains the results from this survey.

---

[3] Genetic sequence data from influenza viruses with human pandemic potential (i.e. viruses that have been found to infect humans and that have a haemagglutinin gene 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).

[4] See PIP Framework Advisory Group, Report to the Director-General, available at http://www.who.int/influenza/pip/pip_ag_oct2014_meetingreport_final_7nov2014.pdf?ua=1.

Methodology Summary

To conduct this survey, the Secretariat developed a questionnaire on the processes and procedures used by data providers and data users in relation to the sequencing, sharing and use of IVPP GSD. The questionnaire focused on the following eight areas, developed on the basis of the technical features identified by the PIP Advisory Group in its October 2014 recommendation to the Director-General[6] and the terms of reference of the PIP Advisory Group's Technical Working Group on the sharing of influenza genetic sequence data[7]: 1) Mechanisms for sharing of IVPP GSD; 2) Ease of sharing; 3) Systematic sharing; 4) Timeliness of sharing; 5) Completeness of data; 6) Traceability; 7) Acknowledgement and collaboration; 8) Quality control.

The questionnaire was sent to data providers and users, including GISRS laboratories, academia and industry. 41 institutions completed the questionnaire: 13 GISRS laboratories, 3 non-GISRS public health laboratories, 12 non-GISRS academic research laboratories, 3 non-GISRS government research laboratories, and 10 manufacturers. More information on the methodology can be found in Annex 1.

[5] In October 2014, the PIP Framework Secretariat also conducted an initial review of the different models of electronic databases that house genetic sequence data of influenza viruses with human pandemic potential ('IVPP GSD').

[6] Meeting of the PIP Framework Advisory Group, 21-24 October 2014, Report to the Director-General: http://www.who.int/influenza/pip/combined_pipagmroct2014corr.pdf?ua=1

[7] TWG on the sharing of influenza genetic sequence data, Terms of Reference: http://www.who.int/influenza/pip/advisory_group/twg_tors.pdf?ua=1

IVPP GSD Sharing in Numbers (as of October 2014)[8]

	GISAID EpiFlu™	INSDC	IRD	OpenFluDB
Number of influenza sequences	400 000 nucleotide sequences of genome segments from approximately 120 000 influenza viruses	381 000 influenza sequences with 350 000 in the specialized Influenza Virus Resource	348 297 segment sequences and 474 700 protein sequences from 82 584 influenza viruses	245 000 sequences in 67 000 influenza viruses
Number of sequences from influenza viruses with human pandemic potential	Sequences of 5712 genome segments of 1051 influenza viruses (isolates/cases) with human pandemic potential (IVHPP); Animal and laboratory derived candidate vaccine viruses (CVV) of relevant subtypes shared with GISRS in the context of the PIP-FW (approximately 30 H5N1 viruses (including 17 CVVs) and 1 H9N2 (CVV)).	About 65 000 including Pandemic (H1N1) 2009 sequences; about 5000 excluding Pandemic (H1N1) 2009 sequences.	3104 sequences from human isolates of H4N8, H7N1, N7N2, H7N3, H7N7, H7N9, H9N2, H10N7, H7N8, excludes all human H1N1, H1N2, H2N2, H3N2, and LAB viruses.	N/A
Number of sequences uploaded into and downloaded monthly from databases	Average of 4856 genetic sequences uploaded to EpiFlu™ on a monthly basis. (period 2010-2014) GISAID has approximately 5500 active users.	At a single site, e.g. GenBank, average monthly upload in the past 12 months: 5500 (range 1200-16 000) and average monthly flu sequence viewed in the past 12 months: 53,000; Average monthly ftp dataset downloads of flu: 7100.	Average of 7340 segment sequence uploads per month into IRD from GenBank (first 9 months of 2014). Average of 3 705 216 sequence downloads per month from IRD (first 9 months of 2014).	Around 2000 sequences uploaded each month.

[8] All data and information in this table was provided by databases in October 2014 in answer to a questionnaire sent by the Secretariat in September 2014. Additional information and data found in Annex 4 of the TWG draft document "Optimal Characteristics of an influenza genetic sequence data sharing system under the PIP Framework" (available at http://www.who.int/entity/influenza/pip/advisory_group/draft_twg_doc.pdf?ua=1).

Survey Results

1. Mechanisms for sharing of IVPP GSD

The PIP Framework sets out several expectations regarding the sharing of IVPP GSD. For instance, section 5.2.1 states that "genetic sequence data, and analyses arising from that data, relating to H5N1 and other influenza viruses with human pandemic potential should be shared in a rapid, timely and systematic manner with the originating laboratory and among WHO GISRS laboratories." In order to gain a better understanding of the mechanisms currently in place for the sharing IVPP GSD in GISRS and non-GISRS laboratories, the Secretariat asked data providers and users how they ***share*** IVPP GSD – whether by using publicly-accessible databases or other means – and how they ***access*** IVPP GSD and why they prefer these mechanisms.

Data providers

The majority of data providers reported uploading IVPP GSD generated by their laboratories to publicly-accessible databases. Only one data provider, from a national public health laboratory, specifically indicated that it does not share IVPP GSD with publicly-accessible databases. This laboratory usually sequences IVPPs to confirm positive cases for their Ministry of Health. Six data providers also indicated that they use other means to share IVPP GSD, including emails and reports. While a majority of data providers use GISAID EpiFlu™, several use more than one database to share IVPP GSD (see Figure 1).

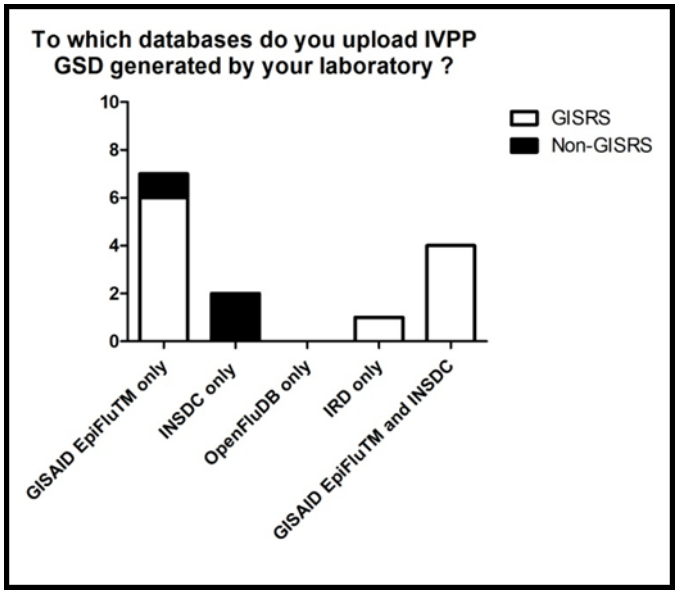

Figure 1. Compiled answers to question 2, part 1, section 1 of Questionnaire (Annex 2)

When asked why they use a specific database to upload data, data providers answered as follows:

	Convenience of data upload and user interface	Data access and use policy	Added value and available tools	Possibility to upload additional data
GISAID EpiFlu™	8/11	8/11	6/11	4/11
INSDC	2/6	2/6	2/6	1/6

Table 1. Compiled answers to question 3, part 1, section 1 of Questionnaire (Annex 2)

Data providers who upload IVPP GSD to GISAID EpiFlu™ choose this database mainly because of convenience of data upload and user interface, and data access and use policy (see Table 1). Several data providers specified that their choice is influenced by certain terms of the data access and usage policy such as the requirements that users register and identify themselves (8/14) and that they acknowledge the originating laboratory/country (8/14). One data provider specified that they upload data to GISAID EpiFlu™ because of its widespread use by GISRS laboratories.

In the case of INSDC, IRD and OpenFluDB, several data providers mentioned that they use these databases because of institutional policies, requirements from scientific journals and out of habit. Comments from data providers indicate that many also choose these databases because they are perceived as offering the broadest possible access to IVPP GSD and they are "widely used by scientists globally".

Only half of GISRS data providers indicated that they have standard operating procedures for uploading IVPP GSD to a publicly-accessible database. One data provider from a GISRS laboratory explained that IVPP GSD were shared "following 1) approval of the sample submitting laboratory, 2) when the data is used in the online WHO VCM summary report, 3) when the virus is selected as a CVV, or 4) following publication in journals".

Data users

The survey indicates that a majority of users access IVPP GSD through more than one database, mainly GISAID EpiFlu™ and INSDC (see Figure 2).

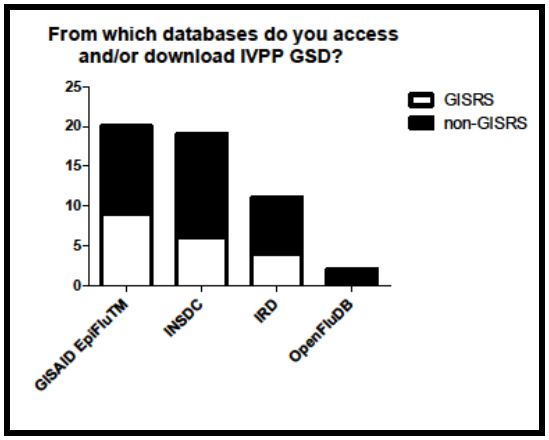

Figure 2. Compiled answers to question 1, part 2, section 2 of Questionnaire (Annex 2)

When asked for which reasons they chose a specific database to access data, users gave the following answers:

	Access to sequences not available in other databases	Convenience of data download and user interface	Data access and use policy	Possibility to access additional data	Added value and available tools
GISAID EpiFlu™	16/20	11/20	12/19	9/20	6/20
INSDC	7/19	12/19	2/6	3/19	3/19
IRD	2/11	9/11	4/11	4/11	5/11
OpenFluDB	0	1/2	2/2	0	0

Table 2. Compiled answers to question 2, part 2, section 2 of Questionnaire (Annex 2)

2. Ease of sharing

The objective of this section was to identify difficulties in sharing or accessing IVPP GSD.

Data providers

All respondents indicated that databases were easy or very easy to use. Although no data providers reported issues related to the upload process, it was mentioned that learning to operate the database requires some training. One data provider also mentioned that GISAID EpiFlu™'s upload tools were easier to use than GenBank's.

Data users

The majority of data users indicated that databases were easy or very easy to use (see Figure 3).

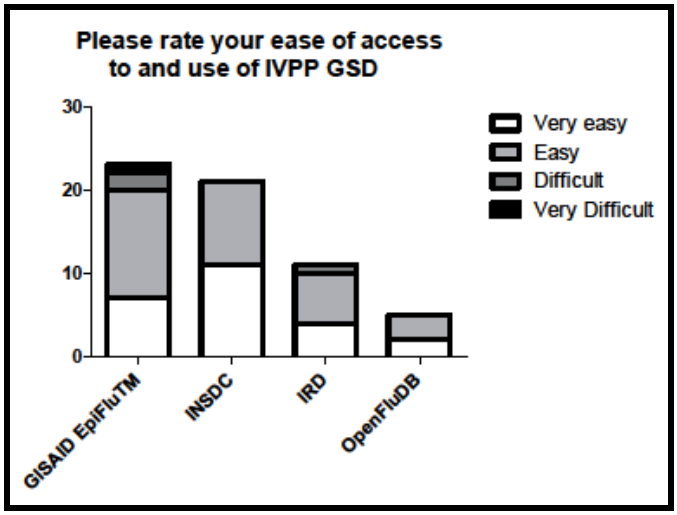

Figure 3 Compiled answers to question 1, part 4, section 2 of Questionnaire (Annex 2)

One user mentioned that they appreciate "GISAID and OpenFluDB for their flexibility in how we can include information in the header of each entry within a FASTA file and parse the data." Two users reported having problems with obtaining access to GISAID EpiFlu™ or obtaining institutional authorization to use it. One data user stated that having to log in to a database was inconvenient, while another data user mentioned that they did not like data access agreements.

3. Systematic sharing

Under the PIP Framework, Member States are encouraged to share biological materials from all cases of IVPP through GISRS, as feasible.[9] The objective of this section was therefore to learn more about which IVPP are sequenced, which IVPP sequences are shared, and why laboratories choose to share these sequences in particular. Users of IVPP GSD were asked whether all relevant sequences needed for their work were available.

[9] PIP Framework Section 5.1.1.

Data providers

Almost all GISRS laboratories sequence all IVPP samples they receive (see Figure 4). One data provider that answered negatively explained that they usually only sequence IVPP samples to confirm positive H5N1 cases for the Ministry of Health.

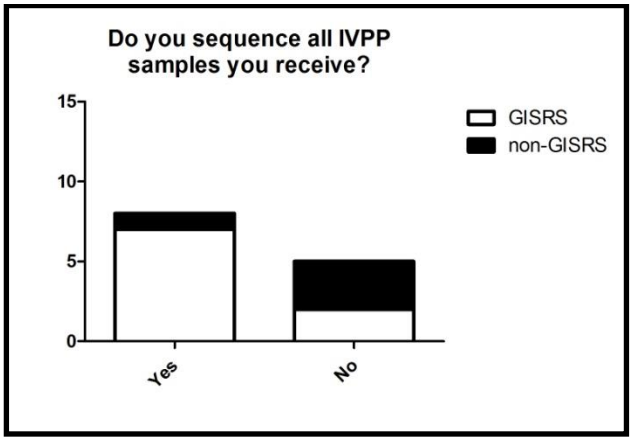

Figure 4. Compiled answers to question 2, part 3, section 1 of Questionnaire (Annex 2)

A total of 10 out of 12 data providers indicated that they share all the IVPP GSD they generate (see Figure 5). One laboratory that did not share all sequences explained that the sequences not shared were quality control sequences of limited interest for outside parties. Another provider specified that, in the past, sequences for certain subtypes were shared, but not all of them.

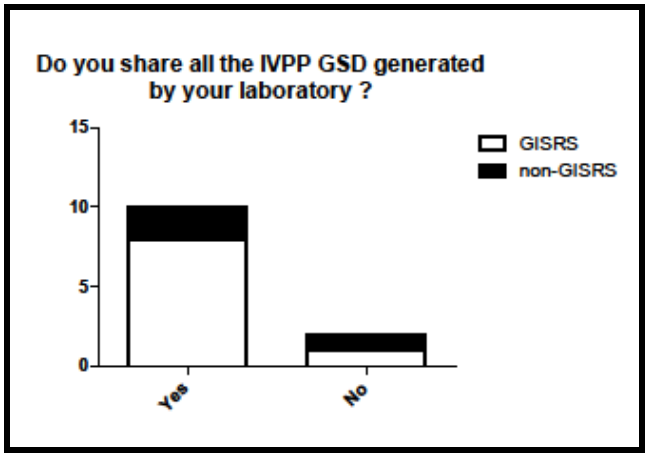

Figure 5. Compiled answers to question 3, part 3, section 1 of Questionnaire (Annex 2)

Out of 18 data providers, 9 indicated that they do not upload all IVPP GSD to publicly-accessible databases. For instance, several GISRS laboratories mentioned that they do not upload IVPP GSD unless they receive permission from the sample-submitting laboratory. We do not know the number of sequences that are not uploaded in publicly-accessible databases because this was not asked in the questionnaire. As providers indicated they share all IVPP GSD they generate, this data is probably shared with the sample-submitting laboratory and/or other GISRS laboratories via email, reports or through other methods.

Data users

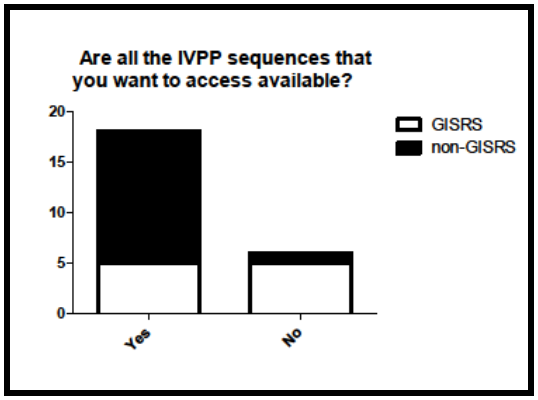

Figure 6. Compiled answers to question 1, part 5, section 2 of Questionnaire (Annex 2)

A majority of users were satisfied that all the IVPP GSD necessary for their work was available (see Figure 6). However, two users pointed out that since they might not know about the existence of certain sequences, it is impossible for them to be sure that they have access to all sequences that might be useful.

One user reported that "cases of H5 and H7 are reported to WHO or during VCM, but the data is not available in databases." A data user from a GISRS laboratory made the following comment on the availability of IVPP GSD: "Many sequences are shared in an exemplary way, particularly through GISAID [EpiFlu™] (e.g. the pandemic H1N1 virus, the first H7N9 virus gene sequences from China.) However, [other centres can be quite slow in depositing certain gene sequences], possibly due to the difference between public access and public domain [databases]."

4. Rapid and timely sharing of IVPP GSD

In order to ensure timely risk assessment and influenza product development, IVPP GSD should be shared rapidly after sample collection and sequencing. The objective of this section was therefore to survey how quickly IVPP GSD are shared and to understand the obstacles (if any) to the rapid sharing of IVPP GSD, and how these obstacles could be addressed.

Data providers

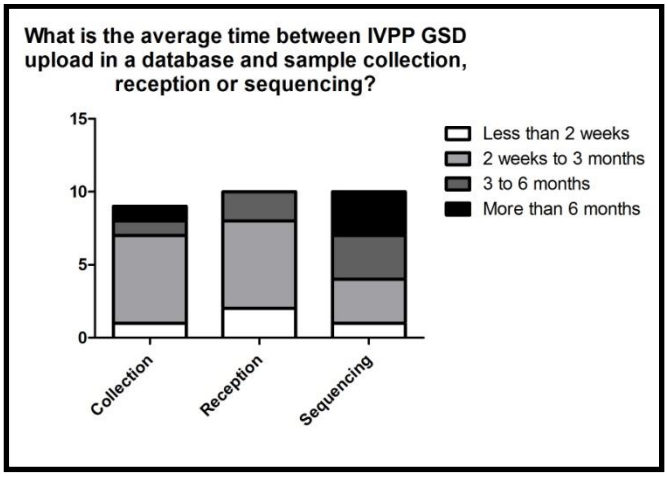

Figure 7. Compiled answers to questions 1-3, part 4, section 1 of Questionnaire (Annex 2)

The majority of providers indicated that IVPP GSD generated for surveillance purposes are usually uploaded in publicly-accessible databases between two weeks and three months after sample collection (see Figure 7, band 1) or reception of the sample by the laboratory (see Figure 7, band 2).

IVPP GSD from samples sequenced for research purposes were uploaded with a longer delay, with only 4 laboratories out of 10 uploading less than three months after sequencing (See Figure 7, band 3).

When asked why they do not upload IVPP GSD immediately after sequencing, three GISRS providers indicated that they need permission from the ministry of health or the originating laboratory before sharing sequences (see Figure 8). One of these laboratories further explained that sharing of IVPP GSD would not occur before notification of the results to WHO and to the originating laboratory. Three laboratories also indicated that delays were due to pending publications or patents (see Figure 8).

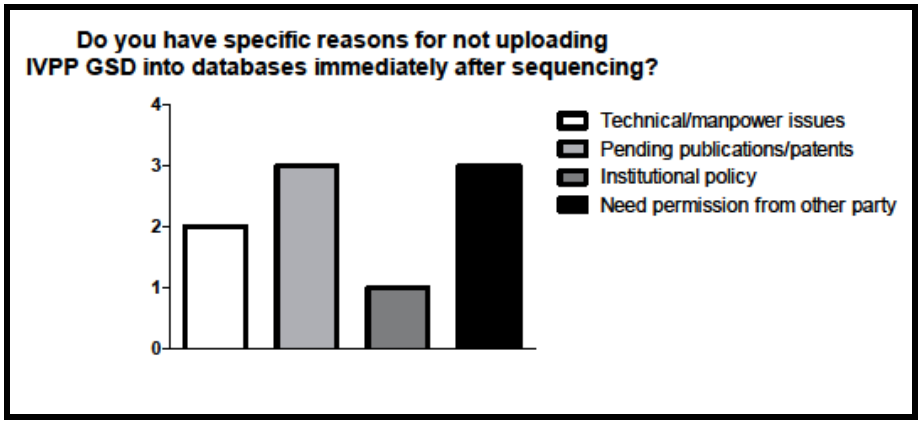

Figure 8. Compiled answers to question 4, part 4, section 1 of Questionnaire (Annex 2)

Data users

Almost all users indicated they needed access to IVPP GSD at least within three months of their upload, with the majority needing access within two weeks, regardless of their intended use for the data (see Figure 9).

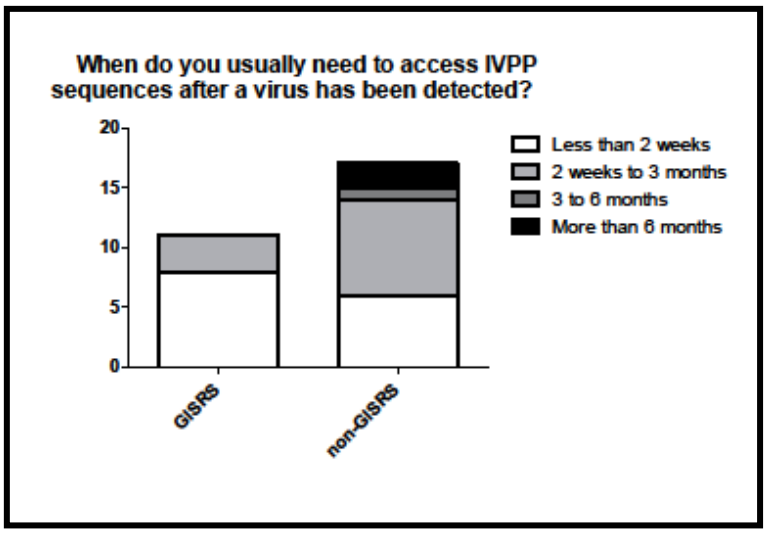

Figure 9. Compiled answers to question 1, part 3, section 2 of Questionnaire (Annex 2)

A total of 16 out of 22 users indicated that they were satisfied with the timeliness of IVPP GSD sharing (see Figure 10). Among the users that reported not having access to the data in a timely manner, three specified that only some of the IVPP GSD were shared too late.

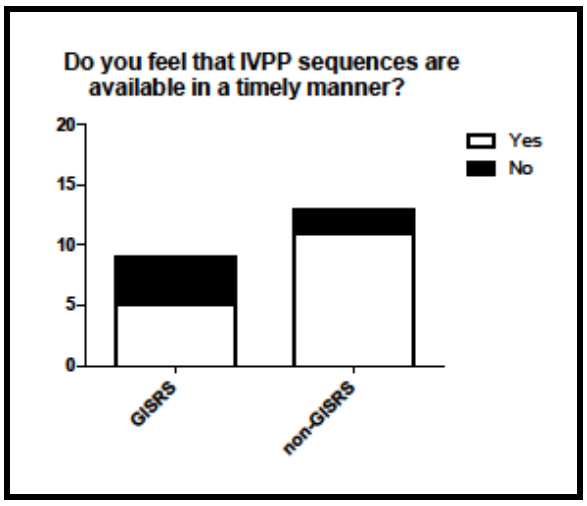

Figure 10. Compiled answers to question 2, part 3, section 2 of Questionnaire (Annex 2)

When asked how they learn about the availability of new IVPP GSD, users indicated that they either conduct frequent manual searches in the databases in order to find new sequences, are informed through WHO/GISRS notifications or, in one case, directly by the database (e-mail notification).

5. Completeness of data

Besides the actual nucleic acid sequence, databases generally allow data providers to upload metadata, which is additional data describing information about the host of the sequenced IVPP sample. In this section, the Secretariat asked data providers and users about access to metadata.

Data providers were asked whether they can find relevant clinical and epidemiological data related to the IVPP samples they sequence. Out of 11 data providers, 9 reported that they have access to the relevant clinical or epidemiological information related to the IVPP samples they sequence (see Figure 11).

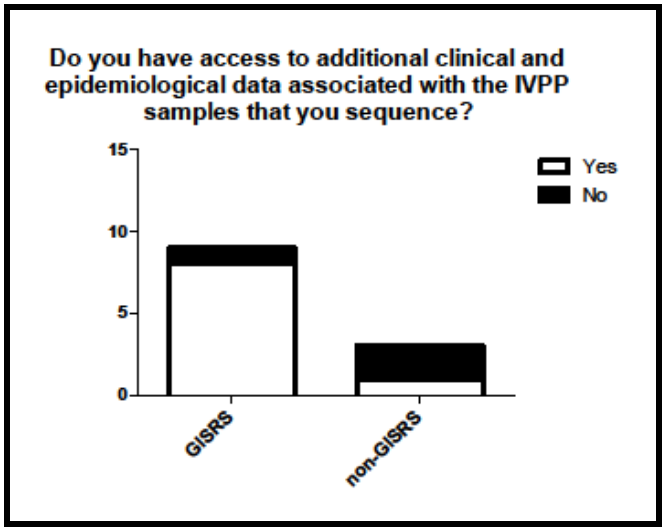

Figure 11. Compiled answers to question 1, part 5, section 1 of Questionnaire (Annex 2)

The majority of data users felt that they could access all the additional data needed for their work (see Figure 12). Users most frequently mentioned that clinical data was missing, with one user listing the following missing clinical data: "patient status (deceased, severe/ICU, mild, asymptomatic); antigenic characterization (HI titres); drug sensitivity (IC50 fold changes for standard drugs); GPS coordinates (for more detailed epidemiological studies)".

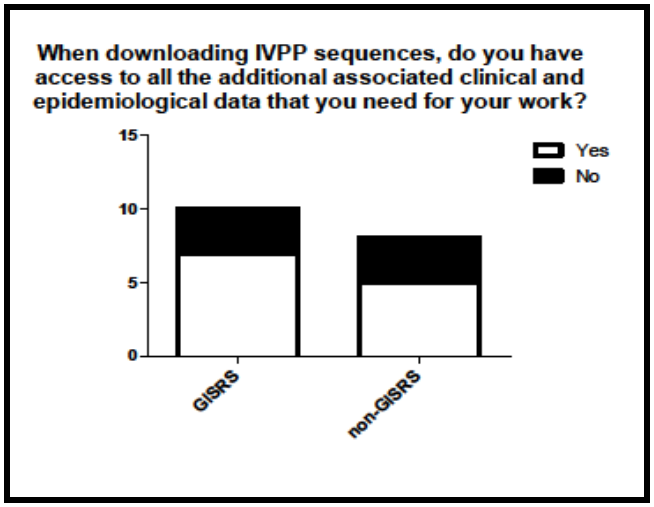

Figure 12. Compiled answers to question 2, part 5, section 2 of Questionnaire (Annex 2)

6. Traceability

Fair and equitable benefit-sharing under the PIP Framework depends on WHO's ability to identify which institutions have used IVPP GSD. Therefore, clear identification of the origin of the viruses and data by laboratories sequencing IVPP and/or using IVPP GSD is required. In this section, data providers and users were asked about practices in place in their institution or company regarding the identification of IVPP GSD.

Data providers

All data providers reported that they specify the country of origin of samples when uploading IVPP GSD. One survey participant reported not identifying the laboratory that performs the original influenza detection, due to not having this information, and one reported not identifying the laboratories performing the sequencing.

Data users

A majority of data users reported that they could find information related to country of origin (21/23) when accessing IVPP GSD. However, information about the sequencing laboratory or the laboratory that isolated the virus is not always easily accessible.

Almost all users answered that they disclose the source of IVPP GSD when publishing results from research projects or clinical trials, mainly using database accession number (see Figure 13).

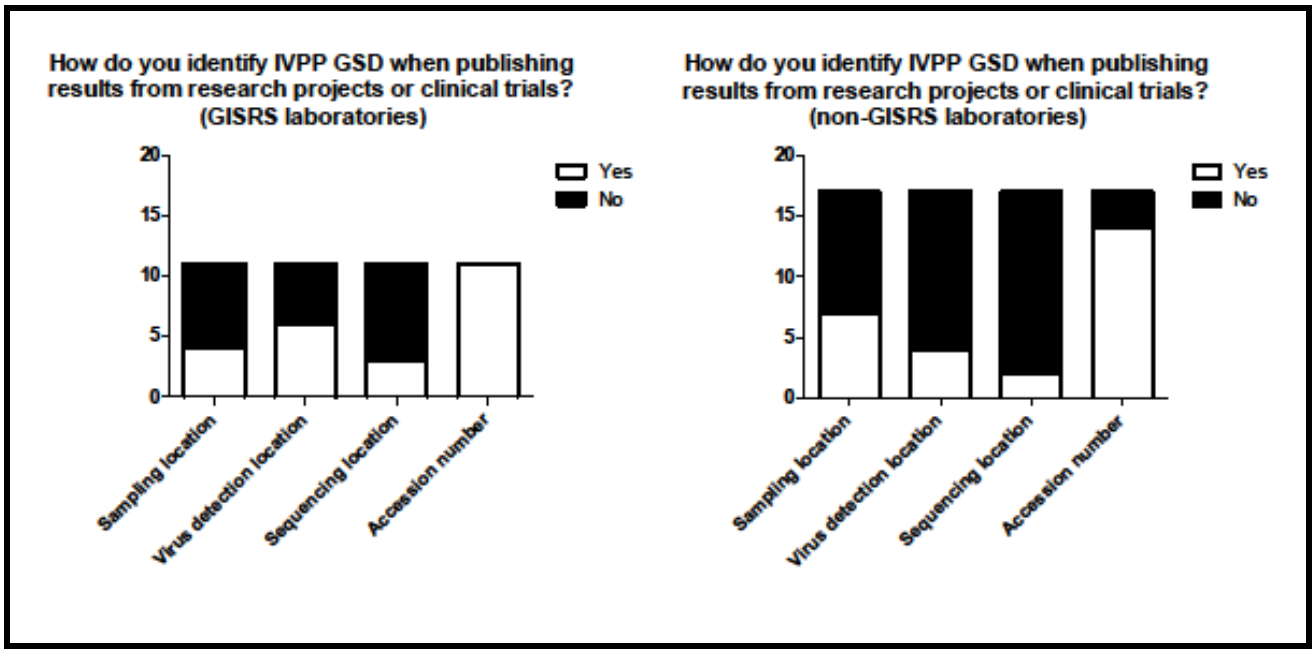

Figure 13. Compiled answers to question 2, part 6, section 2 of Questionnaire (Annex 2).

7. Acknowledgement and collaboration

The PIP Framework encourages the appropriate acknowledgement of the contribution of originating laboratories and collaboration with researchers who generate IVPP GSD. In this section, data providers and users were asked about practices in place in their institution or company related to acknowledgement and collaboration.

Data providers

Only 3 data providers reported being properly acknowledged by users of IVPP GSD or being occasionally contacted for collaboration in research or during the preparation of manuscripts. However, no provider cited this as a potential source of reluctance to share IVPP GSD. One data provider from a GISRS laboratory, for example, said: "I don't think acknowledgement for every single sequence downloaded or used can be expected or is reasonable. Molecular biologists have been working with databases like GenBank for decades where no acknowledgement is required nor usually given (in papers etc). Therefore, I would not expect anybody to acknowledge that they used a sequence from our laboratory for their work."

One data provider and user from a GISRS laboratory, however, explained that "it is essential that those outside of the virology community, for example bioinformaticians, understand the amount of work involved in maintaining the collection of viruses, their characterisation and their archiving. Full acknowledgement of data is essential for the continuation of high quality surveillance of influenza that has evolved over several decades."

Data users

All respondents reported properly acknowledging originating laboratories when using their IVPP GSD. One user from an academic research laboratory, however, explained: "If there are a small number of important sequences then, of course, we acknowledge (or collaborate with) the original data provider. However, if we are performing a large phylogenetic/evolutionary study, then it is really not practical to list everything individually since there would be several thousand sequences." Another user from a GISRS laboratory clarified: "Gene sequences that serve solely as well-established reference sequences are not usually acknowledged specifically but indirectly by accession numbers."

A total of 9 users out of 14 indicated that they sometimes involve originating laboratories in their research. One user mentioned: "Using GISAID, it was fairly easy to contact the data submitter and in several cases it was possible to directly collaborate." However, one user from a GISRS laboratory explained: "Collaborating is and would be extremely difficult in many/most circumstances. It is often hard to come up with an approach where an originating laboratory can contribute significantly to the research. Also, funding agencies do not necessarily support the involvement of originating laboratories. (Note: without extra funding, research participation by another lab may not be possible)."

8. Quality control

This section focused on the procedures used by laboratories to ensure the quality of the IVPP GSD they produce.

Data providers

A total of 7 out of 12 providers indicated they only upload full gene sequences and reported having a formal quality assurance program in place in their laboratory. Only one laboratory from the academic research sector had an automated process for GSD upload.

Data users

Almost all data users reported being satisfied with the quality of the IVPP GSD. One user however specified that “sometimes, not all segments of a new virus are available, non-coding regions are often missing. At other times, it is not clear whether non-coding region sequences are primer-derived or true sequences derived from a virus.”

Annex 1

Methodology

Influenza viruses with human pandemic potential

As the PIP Framework covers only IVPP, this survey focuses on the sharing of GSD from those viruses. Under the Framework, IVPP are "viruses that have been found to infect humans and that have a haemagglutinin gene 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."[10] IVPP are, therefore, understood to include – but are not limited to – the following viruses:

- Human cases of avian influenza viruses e.g. H5N1, H5N6, H5N8, H5N9, H6N1, H7N2, H7N3, H7N7, H7N9, H9N2, H10N7, H10N8
- Human cases of influenza variant viruses e.g. H1N1v, H1N2v, H3N1v, H3N2v

IVPP GSD User/provider survey

a) Survey criteria

The survey focuses on the following eight areas, developed on the basis of the technical features identified by the PIP Advisory Group in its October 2014 recommendation to the Director-General[11] and the terms of reference of the PIP Advisory Group's Technical Working Group on the sharing of influenza genetic sequence data[12]:

1. Mechanisms for sharing IVPP GSD
2. Ease of sharing
3. Systematic sharing
4. Rapid and timely sharing
5. Completeness of data
6. Traceability
7. Acknowledgement and collaboration
8. Quality control

The complete survey can be found in Annex 3.

b) Providers and users of data

For the purpose of this survey, providers of IVPP GSD are laboratories that have sequenced and shared IVPP GSD. These include:

- ***GISRS laboratories***: These are laboratories recognized or designated by the WHO Director-General as a member of GISRS, that sequence and/or upload IVPP GSD as part of their PIP Terms of Reference. GISRS has 4 categories of laboratories: National Influenza Centres ("NIC"), WHO

[10] See PIP Framework, Section 4.2. Available at http://apps.who.int/iris/bitstream/10665/44796/1/9789241503082_eng.pdf
[11] Meeting of the PIP Framework Advisory Group, 21-24 October 2014, Report to the Director-General: http://www.who.int/influenza/pip/combined_pipagmroct2014corr.pdf?ua=1
[12] TWG on the sharing of influenza genetic sequence data, Terms of Reference: http://www.who.int/influenza/pip/advisory_group/twg_tors.pdf?ua=1

Collaborating Centres for Influenza ("WHO CC"), Essential Regulatory Laboratories ("ERL") and H5N1 Reference Laboratories ("H5RL"). The Terms of Reference for each category of laboratory are set out in Annex 5 of the PIP Framework;

- ***Other authorized laboratories***: These laboratories are defined in PIP Framework section 4.3, that may share IVPP with GISRS and produce and/or upload GSD to databases;
- Other public and private laboratories that may sequence PIP biological materials[13] and upload the GSD to databases.

In order to identify as many data providers as possible, a search was conducted in GISAID EpiFlu™ and GenBank of all laboratories that have submitted IVPP GSD since 2011. All laboratories associated to GSD sequences from human isolates of H5N1, H5N6, H5N8, H5N9, H6N1, H7N2, H7N3, H7N7, H7N9, H9N2, H10N7, and H10N8 were listed as providers of IVPP GSD and contacted.

In addition, relevant WHO Collaborating Centres on Influenza, H5 Reference Laboratories and Essential Regulatory Laboratories and National Influenza Centres were included. These were considered as both likely providers and users of IVPP GSD. In total, 61 possible providers were identified and contacted by e-mail (see Annex 2).

Unlike providers of IVPP GSD, who can be identified through publicly-accessible databases, it is difficult to compile an exhaustive list of all users of IVPP GSD. For this survey, we therefore identified the following entities as potential users:

- vaccine manufacturers, influenza diagnostic companies, and other entities involved in influenza research.
- authors of peer-reviewed articles listed in PubMed and containing the terms "influenza", "genetic sequence", and dealing with IVPPs.

c) Results

41 institutions completed the survey: 13 GISRS laboratories, 3 non-GISRS public health laboratories, 12 non-GISRS academic research laboratories, 3 non-GISRS government research laboratories, and 10 manufacturers. Among these, 23 were users only, and 18 were both providers and users.

[13] PIP Framework Section 4.1 defines "PIP biological materials" as such: "for the purposes of this Framework (and its annexed Standard Material Transfer Agreements (SMTAs) and terms of reference (TORs)) and the Influenza Virus Tracking Mechanism (IVTM), includes human clinical specimens, virus isolates of wild type human H5N1 and other influenza viruses with human pandemic potential; and modified viruses prepared from H5N1 and/or other influenza viruses with human pandemic potential developed by WHO GISRS laboratories, these being candidate vaccine viruses generated by reverse genetics and/or high growth re-assortment. Also included in "PIP biological materials" are RNA extracted from wild-type H5N1 and other human influenza viruses with human pandemic potential and cDNA that encompass the entire coding region of one or more viral genes."

Annex 2

Entities Contacted

GISRS entities

Institution	Country
Victorian Infectious Diseases Reference Laboratory, Peter Doherty Institute for Infection & Immunity	Australia
Therapeutic Goods Administration Laboratories	Australia
Institute of Epidemiology Disease Control and Research (IEDCR)	Bangladesh
Instituto Oswaldo Cruz	Brazil
Institut Pasteur in Cambodia	Cambodia
Centre for Health Protection	China
Centre of Influenza Research, The University of Hong Kong	China
Chinese Center for Disease Control and Prevention	China
United States Naval Medical Research Unit No.3	Egypt
Centre National de Référence du Virus Influenza Région Sud	France
Institut Pasteur	France
National Institute of Virology	India
National Institute of Health Research and Development	Indonesia
Istituto Superiore di Sanita	Italy
National Institute of Infectious Diseases	Japan
Institute for Medical Research	Malaysia
Erasmus Medical Centre	Netherlands
Norwegian Institute of Publich Health	Norway
Korea National Institute of Health	Republic of Korea
D.I.Ivanovsky Research Institute of Virology	Russian Federation
Research Institute of Influenza	Russian Federation
State Research Centre for Virology and Biotechnology VECTOR	Russian Federation
National institute for Biological Standards and Control	United Kingdom
The Francis Crick Institute	United Kingdom
Centers for Disease Control and Prevention	United States of America
Food and Drug Administration	United States of America
St. Jude Children's Research Hospital	United States of America
Pasteur Institute	Viet Nam

Non-GISRS entities:

Institution	Country
Sinergium Biotech	Argentina
Australian Animal Health Laboratory - CSIRO	Australia
BioCSL, Ltd.	Australia
School of Botany and Zoology, Australian National University	Australia
Fundacao/Instituto Butantan	Brazil
Center for Disease Modeling, York University	Canada
Department of Biochemistry, Microbiology, and Immunology, University of Ottawa	Canada
Medicago	Canada
Public Health Agency of Canada	Canada
Beijing Institute of Microbiology and Epidemiology	China
Beijing Tiantan Biological Products	China
Changchun Changsheng Life Sciences	China
Changchun Changsheng Life Sciences Ltd	China
Changsha CDC	China
Chengdu Kanghua Biological Products Co.,Ltd	China
China National Biotec Group Company Ltd	China
Chinese Center for Animal Health and Epidemiology	China
College of Animal Science and Medicine, Inner Mongolia Agriculture University	China
College of Medicine, Zhejiang University	China
Department of Biological Sciences and Biotechnology, Zhangzhou Normal University	China
Fujian CDC	China
Guangdong Provincial CDC	China
Guangzhou CDC	China
Hangzhou CDC	China
Harbin Veterinary Research Institute	China
Hualan Biological Engineering	China
Hunan Provincial CDC	China
Institute of Life Sciences, Taishan Medical College	China
Institute of Pathogen Biology, Chinese Academy of Medical Sciences and Peking Union Medical College	China
Institute Pasteur of Shanghai	China
Jiangsu Provincial CDC	China
Key Laboratory of Systems Biology of Pathogens	China
Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.	China
Laboratory of Animal Virology, Huazhong Agricultural University	China
Laboratory of Microbiology, China Agricultural University	China

Lishui CDC	China
Nanchang CDC	China
Nanjing Medical University	China
Queen Mary Hospital, The University of Hong Kong	China
School of Biological Sciences, The University of Hong Kong	China
Shanghai Institute of Biological Products Co, Ltd	China
Shanghai Municipal CDC	China
Shanghai Public Health Clinical Center affiliated to Fudan University	China
Shanghai Zhijiang Biotechnology Co.,Ltd	China
Sinovac Biotech Ltd	China
State Key Laboratory for Agrobiotechnology	China
State Key Laboratory of Diagnosis and Treatment of Infectious Diseases	China
State Key Laboratory of Emerging Infectious Diseases	China
State Key Laboratory of Genetic Engineering	China
State Key Laboratory of Pathogen and Biosecurity	China
State Key Laboratory of Respiratory Diseases	China
The School of Preclinical Medicine, Sun Yatsen University	China
Wuhan Institute of Virology	China
Xiangxue group Co., Ltd	China
Yunnan Walvax Biotech Co.	China
Zhejiang Provincial CDC	China
Research Institute of Tropical Roots, Tuber Crops and Plantains	Cuba
The Panum Institute	Denmark
Virus Research and Development, Statens Serum Institute	Denmark
Faculty of Veterinary Medicine, Beni-Suef University	Egypt
National Research Centre	Egypt
Fabentech	France
Sanofi S.A.	France
Euroimmun	Germany
IDT Biologika GmbH	Germany
Institute of Medical Virology, Justus Liebig University	Germany
Institute of Molecular Virology, Universitaetsklinikum Muenster	Germany
Institute of Virology, Philipps University	Germany
Institute of Virology, University Medical Center Freiburg	Germany
Westfaelische Wilhelms-Universitet Muenster	Germany
Omninvest	Hungary
Department of Applied Mathematics and Humanities, S. V. National Institute of Technology	India
IILM-AHL	India
International Centre for Genetic Engineering & Biotechnology	India
Tasinim Biotechnology Research Center	Iran
The Conway Institute of Biomolecular and Biomedical Research	Ireland

Division of Avian and Aquatic Diseases, Kimron Veterinary Institute	Israel
Department of Infectious, Parasitic and Immune-Mediated Diseases, Istituto Superiore di Sanità	Italy
Istituto Zooprofilattico Sperimentale delle Venezie	Italy
Department of Infectious Medicine, Kurume University School of Medicine	Japan
Institute of Medical Science, University of Tokyo	Japan
Kaketsuken,The Chemo-Sero-Therapeutic Research Institute	Japan
Kitasato Daiichi Sankyo Vaccine Co., Ltd	Japan
Nagahama Institute of Bio-Science and Technology	Japan
Research Institute for Microbial Diseases	Japan
Departamento de Medicina y Zootecnia de Aves, Universidad Nacional Autónoma de México	Mexico
National Institute for Public Health and the Environment	Netherlands
Celltrion, Inc.	Republic of Korea
Chungnam National University	Republic of Korea
Green Cross Corporation	Republic of Korea
Il-Yang, Korea	Republic of Korea
Korea Institute of Science and Technology Information	Republic of Korea
SK Chemicals	Republic of Korea
South Korea Konkuk University	Republic of Korea
Microgen	Russian Federation
Microgen	Russian Federation
St Petersburg Institute of Vaccine and Sera	Russian Federation
Bioinformatics Institute, Agency for Science, Technology and Research	Singapore
Department of Biochemistry, National University of Singapore	Singapore
Duke-NUS Graduate Medical School, Singapore Center for Infectious Diseases	Singapore
Genome Institute of Singapore	Singapore
Temasek Life Sciences Laboratory	Singapore
Vircell SL	Spain
The Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences	Sweden
Institute of Virology and Immunoprophylaxis	Switzerland
Novartis International AG	Switzerland
Swiss Institute of Bioinformatics	Switzerland
Bureau of Epidemiology, Ministry of Public Health	Thailand
Chulalongkorn University	Thailand
Department of Immunology, Mahidol University	Thailand
Department of Microbiology, Mahidol University	Thailand
GPO-Merieux Biologicals Ltd	Thailand
Central Veterinary Laboratory-Weybridge	United Kingdom

Department of Pathology, University of Cambridge	United Kingdom
Department of Virology, Imperial College London	United Kingdom
Department of Zoology, University of Cambridge	United Kingdom
Department of Zoology, University of Oxford	United Kingdom
Division of Infectious Disease, Imperial College London	United Kingdom
GlaxoSmithKline plc	United Kingdom
Institute of Evolutionary Biology, University of Edinburgh	United Kingdom
Sir William Dunn School of Pathology, University of Oxford	United Kingdom
Alere Scarborough, Inc.	United States
BioFire Diagnostics	United States
Computer Science Department and Lane Center for Computational Biology, Carnegie Mellon University	United States
Department of Ecology and Evolutionary Biology, University of California-Irvine	United States
Department of Ecology and Evolutionary Biology, University of Colorado	United States
Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor	United States
Department of Microbiology, Mount Sinai School of Medicine	United States
Department of Pathology, University of California, San Diego	United States
Department of Pathology, University of Texas Southwestern Medical Center	United States
Department of Veterinary Microbiology and Preventive Medicine, Iowa State University	United States
Diatherix Laboratories Inc.	United States
Focus Diagnostics, Inc.	United States
Hemispherx Biopharma, Inc.	United States
IndevR Inc	United States
Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin	United States
IntelligentMDx, Inc.	United States
J Craig Venter Institute	United States
Laboratory of Infectious Diseases, NIAID/NIH	United States
Massachussets Institute of Technology	United States
MedImmune	United States
Nanotherapeutics, Inc.	United States
Newport Laboratories	United States
Novavax	United States
Odum School of Ecology, University of Georgia, Athens	United States
PaxVax Inc.	United States
Penn State Dickinson School of Law	United States
Princeton Biomeditech Corp	United States
Protein Sciences Corporation	United States

Quidel Corporation	United States
SA Scientific, Ltd	United States
Southeast Poultry Research Laboratory	United States
Technovax, Inc	United States
The Southeastern Cooperative Wildlife Disease Study, University of Georgia, Athens	United States
University of Pittsburgh School of Medicine	United States
University of Rochester	United States
USDA Agriculture Research Service	United States
Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center	United States
VaxInnate Corporation	United States
Virapur LLC	United States
Oxford University Clinical Research Unit	Viet Nam
Vabiotech	Viet Nam

Annex 3

SURVEY ON THE SHARING OF GENETIC SEQUENCE DATA OF INFLUENZA VIRUSES WITH HUMAN PANDEMIC POTENTIAL

QUESTIONNAIRE FOR DATA PROVIDERS AND USERS

Background

The PIP Framework is an international arrangement, adopted in 2011 by the 194 Member States of the World Health Organization (WHO), that aims to improve global pandemic influenza preparedness and response, by:

i) improving and strengthening the sharing of influenza viruses with human pandemic potential[14] ('IVPP') through a WHO-coordinated network of influenza laboratories (known as 'GISRS'), and;

ii) promoting the fair and equitable access, by developing countries, to the benefits arising from such sharing.

Genetic sequence data and the PIP Framework

During PIP Framework negotiations, Member States recognized the importance of genetic sequence data for pandemic preparedness and response and requested that the Director-General seek advice from the PIP Advisory Group ('PIP AG') on the "best process for further discussion and resolution of issues relating to the handling of genetic sequence data from H5N1 and other [IVPPs] as part of the Pandemic Influenza Preparedness Framework."

The matter has gained importance given the recent development of synthetic biology technologies which allow the production of influenza virus proteins, antibodies and candidate vaccine viruses using only genetic sequence data. These developments raise questions about the broader implications of sharing and using IVPP GSD, notably with respect to benefit sharing under the PIP Framework.

PIP Advisory Group Guidance on the best process for further discussion and resolution of the issues relating to the handling of GSD

In its October 2014 meeting report to the Director-General[15], the PIP AG made the following observations:

> "a. Laboratories should continue to share [IVPP GSD] as soon as it becomes available because it is necessary for timely and comprehensive pandemic risk assessment and response.
>
> […]

[14] Genetic sequence data from influenza viruses with human pandemic potential (i.e. viruses that have been found to infect humans and that have a haemagglutinin gene 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.

[15] See PIP Framework Advisory Group, Report to the Director-General , available at http://www.who.int/influenza/pip/pip_ag_oct2014_meetingreport_final_7nov2014.pdf?ua=1.

c. The objective of benefit-sharing may be met by mechanisms related to monitoring products generated using influenza GSD, rather than by monitoring use of GSD and/or tracing GSD, noting that source identification is critical.

d. Closer collaboration regarding open sharing of influenza GSD among the many different databases is desirable."

Thus, in its guidance to the Director-General, the PIP AG recommended a process to identify

> "the optimal characteristics of a system for the handling of IVPP GSD under the Framework, including consideration of data sharing systems that are best suited to meet the objectives of the Framework considering obligations and timeliness of data submission, quality assurance of data, completeness of data annotation, ease of access to data, sustainability and security of the system".

Review of data sharing systems

The first step of this process will be to conduct a review of current systems in place for the sharing of influenza genetic sequence data.[16] This information will help the *Technical Working Group (TWG) on the sharing of influenza genetic sequence data* to develop a document defining the optimal characteristics of a GSD sharing system that is best suited to meet the objectives of the Framework and best practices for operationalizing such a system.

Purpose of this questionnaire

To this end, the Secretariat developed a questionnaire for data providers and users, including GISRS laboratories, academia and industry, in order to gain a better understanding of how IVPP GSD is generated, shared and used. As a provider or user of data, your input is invaluable to this process.

Instructions

Your institution was selected because, from 2011 to 2015, you have either:

- sequenced IVPPs (i.e. you are a data provider) or;
- used IVPP GSD to conduct risk assessments, research or to develop pandemic influenza product (i.e. you are a data user).

Therefore, this questionnaire is separated in 2 Sections, which each have 8 parts. Depending on the work you do, WHO would greatly appreciate if you could answer questions from one or both of these Sections:

- Section 1 contains questions for institutions that sequence influenza viruses. Please complete this section only if your institution is a data provider.
- Section 2 contains questions for institutions that use IVPP GSD. Please complete this section only if your institution is a data user.

[16] In October 2014, the PIP Framework Secretariat also conducted an initial review of the different models of electronic databases that house genetic sequence data of influenza viruses with human pandemic potential ('IVPP GSD').

If your institution is both a data provider and a data user, WHO would be extremely grateful if you answered both sections.

Please note that your answers will not be released or shared by WHO with any company, institution, establishment or other entity. The findings from the questionnaire will be anonymized, consolidated and summarized in a report that will be provided to the TWG and the PIP Advisory Group (PIP AG). This report will be shared with you.

For any questions, please contact the PIP framework secretariat at pipframework@who.int.

About your institution

Name of Institution:

Address:

Type of institution (GISRS laboratory, academic research institution, public health laboratory, manufacturer, veterinary laboratory,…)

Type of influenza work conducted:

Details of the person completing this questionnaire:

-Name:

-Job title:

-Email:

SECTION 1. FOR INSTITUTIONS THAT SEQUENCE INFLUENZA VIRUSES WITH HUMAN PANDEMIC POTENTIAL ('IVPP")

Under the Framework, IVPPs are viruses that have been found to infect humans and that have a haemagglutinin gene 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. For ease of reference, IVPPs are understood to include – but are not limited to – the following viruses:

- Human cases of avian influenza viruses e.g. H5N1, H5N6, H5N8, H5N9, H6N1, H7N2, H7N3, H7N7, H7N9, H9N2, H10N7, H10N8
- Human cases of influenza variant viruses e.g. H1N1v, H1N2v, H3N1v, H3N2v

WHO would like to understand how institutions that sequence IVPPs share the GSD they generate. This section has 8 parts, which each contains several questions on the procedures and practices currently in place in laboratories regarding the generation of IVPP GSD and its sharing.

Please answer the questions below if your laboratory has sequenced IVPPs between 2011 and 2015. If you have not, you may go directly to Section II.

Part 1 - Sharing of IVPP GSD: *WHO would like to gain a better understanding of the processes and procedures in place for sharing IVPP GSD, including which databases are used by laboratories to share IVPP GSD and to understand why they share their data using these databases.*

1. Does your laboratory have standard procedures regarding the **uploading** of IVPP GSD into external databases? Yes/No
    a. If yes, please describe the procedures:

2. Do you upload IVPP GSD generated by your laboratory to **publicly-accessible databases**? Yes/No
    a. If yes, please specify in which databases. (Multiple ticks possible):
        - INSDC (GenBank, DDBJ or EMBL-EBIL)
        - GISAID EPIFLU™
        - OpenFluDB
        - If other, please specify:

3.
    a. For which reasons do you upload IVPP GSD to these databases?

	INSDC (DDBJ, EMBL-EBI, GenBank)	GISAID Epiflu™	OpenFluDB	Other:_____
For their data access and use policy				
For convenience of data upload and user interface				

Because of the added value/tools available on these databases				
Because of institutional policy or guidelines				
Because they allow you to upload other relevant data than GSD				
Other reasons, Please explain:				

b. If you have chosen "For their data access and use policy" in the previous question, please specify for which reason(s) (Multiple ticks possible):
   - Users are required to register and identify themselves;
   - There are conditions for further sharing of downloaded data with a third-party;
   - There are conditions for uploading downloaded data to another database;
   - The policy requires acknowledgment of originating laboratory/country;
   - The policy requires collaboration with originating laboratory/country;
   - The policy addresses intellectual property rights and other restrictions on the data;

4. Do you use other means to share IVPP GSD? Yes/No
   a. Please specify which ones:

Please add any other comments related to Part 1 "Sharing of IVPP GSD":

Part 2 - Ease of sharing: *In this part, WHO would like to know if you have any difficulty sharing IVPP GSD.*

Please answer this question in relation to the database you use most regularly.

1. Which tools or systems do you use to upload IVPP GSD to the database?

2. Please rate the ease of use of these tools or systems.

Very easy	Easy	Difficult	Very difficult

a. Please explain.

3. Have you encountered any issues in relation to uploading IVPP GSD to a database? Yes/No

a. If yes, please explain.

b. How do you think these issues could be addressed?

Please add any other comments related to Part 2 "Ease of sharing ":

Part 3 - Systematic sharing: *The Framework encourages Member States to share all cases of IVPPs through GISRS, as feasible. Therefore, in part 3, WHO would like to learn more about which IVPPs are sequenced, which IVPP sequences are shared and why laboratories choose to share these sequences in particular.*

1. From where do you obtain the IVPP samples that you sequence (Multiple ticks possible):
   - From hospital/primary care provider
   - From GISRS lab
   - If other origin, please specify:

2. Do you sequence all IVPP samples you receive? Yes/No
   a. If no, please specify which IVPP you sequence:

   b. Please explain the criteria you apply to select which viruses to sequence:

3. Do you share all IVPP GSD generated by your laboratory? Yes/No

a. If no, please identify why (Multiple ticks possible):
   - National laws and regulations
   - Institutional policy or guideline
   - Sequences are identical or nearly identical to previously uploaded sequences
   - Patents or publications pending
   - If other reasons, please specify:

4. Provide information on the number of sequences of each type/subtype of IVPP shared since 2011.

Virus	Full genome sequences:	HA gene sequences:	NA gene sequences:
H5NX			
H7NX			
Others:			

Please add any other comments related to Part 3 "Systematic sharing ":

Part 4 - Rapid and timely sharing of IVPP GSD: *Timely risk assessment and influenza product development require that IVPP GSD be shared rapidly after sample collection and sequencing. WHO would like to know how quickly IVPP GSD are shared and to understand the obstacles (if any) to the rapid sharing of IVPP GSD, and how these obstacles can be addressed.*

1. If you perform virus typing/subtyping using sequencing, what is the average time between sample collection and IVPP GSD upload in a database?
   a. Less than 2 weeks
   b. 2 weeks to 3 months
   c. 3 to 6 months
   d. More than 6 months

2. If you perform virus typing/subtyping using sequencing, what is the average time between reception of the sample in your laboratory and IVPP GSD upload in a database?
   a. Less than 2 weeks
   b. 2 weeks to 3 months
   c. 3 to 6 months
   d. More than 6 months

3. If you perform sequencing for research purposes, what is the average time between sequencing of IVPP samples and IVPP GSD upload in a database?
   a. Less than 2 weeks
   b. 2 weeks to 3 months

c. 3 to 6 months
d. More than 6 months

4. Do you have specific reasons for not uploading IVPP GSD into databases immediately after sequencing? Yes/No
   - National laws and regulations
   - Institutional policy or guideline
   - technical/manpower issues
   - pending publications or patents
   - Other:

5. If you answered "yes" to question 3 and identified "technical/manpower issues", how could these issues be improved in order to speed up the process of sharing IVPP GSD?

Please add any other comments related to Part 4 "Rapid and timely sharing of IVPP GSD":

Part 5 - Completeness of data: *Genetic sequence databases generally allow data providers to include additional information (including clinical and epidemiological data) when uploading GSD. WHO would like to know if, from your perspective, the databases you use allow you to share all additional information about the IVPP GSD you uploaded.*

Please answer this question in relation to the databases you use most regularly.

1. Do you have access to additional clinical and epidemiological data associated with the IVPP samples that you sequence? Yes/No

   a. If no, please identify missing information (Multiple ticks possible):
   - Date of sample collection
   - Date of symptom onset
   - Age
   - Gender
   - Health status
   - Antiviral treatment
   - Vaccination details

- If other information, please specify:

2. Are you able to include this information when uploading IVPP GSD to a database? Yes/No
    a. If no, please specify why. (Multiple ticks possible)
        - Data unavailable
        - Database does not collect this information
        - Institutional policy or national laws and regulations prevent it
        - Technical/manpower issues
        - If other reason, please explain:

Please add any other comments related to Part 5 "Completeness of data":

Part 6 - Traceability: *In order to implement fair and equitable benefit-sharing under the PIP Framework, WHO may need to identify which institutions have used the IVPP GSD. This would require that laboratories sequencing IVPPs identify clearly the origin of the data and viruses. In this part, WHO would like to know what is the practice in place in your institution regarding identification of IVPP GSD.*

1. When uploading IVPP GSD, do you indicate:
    - The country of origin of the sample? Yes/No
    - The name and contact information of the laboratory that performed the virus detection? Yes/No
    - The name and contact information of the laboratory that performed the sequencing? Yes/No

2. If you do not include the origin of the data, what is the reason? (Multiple ticks possible)
    - Data not available
    - Database does not collect these data
    - Other:

Please add any other comments related to Part 6 "Traceability":

Part 7 - Acknowledgement and collaboration: *The PIP Framework encourages laboratories that use IVPPs*

to appropriately acknowledge the contribution of the originating laboratory and to collaborate with researchers who produced the IVPP GSD.

1. When other laboratories use IVPP GSD sequenced by your laboratory, do you feel that they acknowledge appropriately your laboratory as the source of the data appropriately in publications and presentations? Yes/No

    a. If no, please explain:

    b. If no, does this make you feel reluctant to share IVPP GSD ? Yes/No
        - Please explain:

2. When other laboratories use IVPP GSD sequenced by your laboratory, do they contact your laboratory to collaborate on research and involve you in the preparation of manuscript for publications? Yes/No

    a. If no, please explain:

    b. If no, does this make you reluctant to share IVPP GSD ? Yes/No
        - Please explain:

Please add any other comments related to Part 7 “Acknowledgement and collaboration”:

Part 8 - Quality control: *In this part, WHO would like to determine how laboratories ensure the quality of the IVPP GSD that they produce.*

1. Do you upload full gene sequences only? Yes/No

2. Do you have a quality assurance program in place in your laboratory? Yes/No

    a. If yes, please identify which standards it is based on : (Multiple ticks possible)
        - ISO 15189
        - Standards set by national accreditation agency
        - If others, please identify:

    b. If yes, does your quality assurance program include:
        - standards for identification and labelling of sample? Yes/No
        - standards for controls? Yes/No
        - standards for calibration and maintenance of equipment? Yes/No
        - standards for procedures and policies for nonconformities detection and report? Yes/No

3. Is the process to upload IVPP GSD performed manually or automatically in your laboratory? Manually/Automatically

Please add any other comments related to Part 8 "Quality control":

Please add any other comments related to the sharing of IVPP genetic sequence data:

SECTION 2: FOR INSTITUTIONS THAT USE GENETIC SEQUENCE DATA ('GSD') FROM INFLUENZA VIRUSES WITH HUMAN PANDEMIC POTENTIAL ('IVPP')

Under the Framework, IVPPs are viruses that have been found to infect humans and that have a haemagglutinin gene 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. For ease of reference, IVPPs are understood to include – but are not limited to – the following viruses:

- Human cases of avian influenza viruses: H5N1, H5N6, H5N8, H5N9, H6N1, H7N2, H7N3, H7N7, H7N9, H9N2, H10N7, H10N8
- Human cases of swine influenza variants: H1N1v, H1N2v, H3N1v, H3N2v

WHO would like to understand how institutions access and use IVPP GSD. This section has 8 parts, which each contains several questions on the procedures and practices currently in place in laboratories regarding the use of IVPP GSD

Please answer the questions below if your laboratory has used IVPP GSD between 2011 and 2015.

Part 1 - Use of IVPP GSD: *WHO would like to learn more about how your institution uses IVPP GSD.*

1. For which purpose do you use IVPP GSD? (Multiple ticks possible)
   - Risk assessment
   - Research
   - Development and manufacture of influenza-related products (ex: vaccines, diagnostics,…)
   - Other purposes, please explain:

Please add any other comments related to Part 1 "Use of IVPP GSD":

Part 2 - Access to IVPP GSD: *WHO would like to learn more about how you access IVPP GSD.*

1. From which databases do you access and/or download IVPP GSD?
   - INSDC (GenBank, DDBJ or EMBL-EBI)
   - IRD: Influenza Research Database
   - GISAID EPIFLU™
   - OpenFluDB

- If other, please specify:

2. For which reasons do you use these databases?

	INSDC (GenBank, DDBJ, EMBL-EBI)	GISAID Epiflu™	OpenFluDB	IRD:_____	Other:_____
For their data access and use policy					
For convenience of data download and user interface					
Because of the added value/tools available on these databases					
Because they contain sequences not available in other databases					
Because they allow you to access other relevant data in addition to GSD					
Other reasons, please explain:					

Please add any other comments related to Part 2 "Access to IVPP GSD":

Part 3 - Rapid and timely access to IVPP GSD: *WHO would like to know if IVPP GSD is available when you need it.*

1. When do you usually need to access IVPP sequences after a virus has been detected? (Check one)
   a. Less than 2 weeks
   b. 2 weeks to 3 months
   c. 3 to 6 months
   d. More than 6 months

2. Do you feel that IVPP sequences are available in a timely manner? Yes/No
   a. If no, please explain:

3. How are you informed of the availability of new sequences uploaded in the databases? Please explain.

Please add any other comments related to Part 3 "Rapid and timely access to IVPP GSD":

Part 4 - Ease of access and use of data: *In this part, WHO would like to know if you have any difficulty accessing IVPP GSD.*

Please answer this question in relation to the databases you use most regularly.

1. Please rate your ease of access to and use of IVPP GSD.

	Very easy	Easy	Difficult	Very difficult
INSDC (GenBank, DDBJ, EMBL-EBI)				
GISAID Epiflu™				
OpenFluDB				
IRD				
Other				

   a. Please explain.

2. Have you encountered any issues in relations to access to a database? Yes/No

   a. If yes, please explain.

   b. How do you think these issues should be addressed?

Please add any other comments related to Part 4 "Ease of access and use of data":

Part 5 - Completeness of data: *WHO would like to know whether you have access to all the IVPP sequences and relevant clinical and epidemiological data needed for your work.*

1. Are all the IVPP sequences that you want to access available? Yes/No
   a. If no, please explain:

2. When downloading IVPP sequences, do you have access to all the relevant associated clinical and epidemiological data? Yes/No
   a. If no, what data are missing?:

Please add any other comments related to Part 5 "Completeness of data":

Part 6 - Traceability: *In order to implement fair and equitable benefit-sharing under the PIP Framework, WHO may need to identify which institutions have used the IVPP GSD. This would require that laboratories using IVPPs GSD identify clearly the origin of the data they have used in their work. In this part, WHO would like to know what is the practice in place in your institution regarding identification of IVPP GSD.*

1. When accessing IVPP GSD, can you always find:
   - The country of origin of the sample? Yes/No
   - The name and contact information of the laboratory that isolated the virus? Yes/No

- The name and contact information of the laboratory that performed the sequencing? Yes/No

2. When publishing results from research projects or clinical trials, do you disclose the origin of data you used for research and/or product development? Yes/No
   b. How do you identify the data? (Multiple ticks possible)
      - Sampling location
      - Virus detection location
      - Sequencing location
      - Database Accession number
      - Other information:

Please add any other comments related to Part 6 "Traceability":

Part 7 - Acknowledgement and collaboration: *The PIP Framework encourages laboratories that use IVPPs to appropriately acknowledge the contribution of the originating laboratory and to collaborate with researchers who produced the IVPP GSD.*

1. When using IVPP GSD generated by other laboratories, do you acknowledge these laboratories? Yes/No
   a. If yes, please explain how:

   b. If no, please explain why:

2. When using IVPP GSD generated by other laboratories, do you actively involve researchers from these laboratories in your research projects ?

Please add any other comments related to Part 7 "Acknowledgement and collaboration":

Part 8: Quality control: *WHO would like to identify issues with the quality of the IVPP GSD that you access.*

1. Are you satisfied with the quality of the IVPP GSD that you access? Yes/No
   a. If no, please explain:

Please add any other comments related to Part 8 “Quality control”:

Please add any other comments related to the sharing of IVPP genetic sequence data: