Options to monitor the use of genetic sequence data from influenza viruses with human pandemic potential (IVPP GSD) in end-products

General comments

1. The focus of the draft options document is primarily on monitoring end-products (‘downstream’ options) leaving a discussion of possible ‘upstream’ options for tracing and monitoring the downloading of GSD from databases almost entirely “off the table”.

   This is very unfortunate because it is likely that methods to monitor both GSD downloads and end products will need to be put in place in order to accurately monitor use of IVPP GSD to meet the objectives of the PIP Framework. In order to evaluate the merits of all available options for monitoring use of GSD objectively, we must have a clearer description of the features of existing publicly accessible databases, with and without “access agreements and registration requirements”, in relation to conditions for use of data. This is essential to understanding how conditions may be ‘imposed’ on the use of data and the feasibility of attaching any obligations on the use of data for end-products, under the PIP Framework, particularly after the data have been provided to a public-domain database.

2. The document has substituted the clear and definite language offered by the PIP Framework with terminology that is confusing. The authors should refrain from introducing novel terms that are at best ambiguous. For example the term “public open access database” introduced in this document to refer to “databases that allow free access to the data they house without any conditions on the access, use or distribution of data”, is not used in the PIP Framework, nor in previous related documents, and is quite unhelpful. More consistent use of the single term ‘publicly-accessible’ or use of the clear distinction between “public-domain and public-access databases such as GenBank and GISAID, respectively” is necessary to assist appreciation of the feasibility of imposing conditions on the use of IVPP data in the options proposed.¹

3. While on the one hand, all databases do have Terms of Use, placing GSD into an archive like GenBank (operated in the United States and therefore subject to the US legal definition for information placed in the Public Domain) relinquishes any inherent (Intellectual Property) rights. The DNA Data Bank of Japan (DDBJ) also makes clear that “The INSDC will not attach statements to records that restrict access to the data, limit the use of information in these records, or prohibit certain types of publications based on these records.”² The draft document therefore does not fully acknowledge the difficulties faced by the PIP-FW in attaching obligations (e.g. Partnership Contributions) to the use of such data.
4. While stated that “IVPP are part of a broader set of materials called --- ‘PIP BM’”, IVPP GSD are not defined in that context within the PIP Framework, since IVPP GSD include only viruses “found to infect humans” (section 4.2), but not, for example, viruses from animals that have been shared with GISRS and used to develop candidate vaccine viruses (included in PIP BM). While GSD of such viruses would need to be included under a revised definition of IVPP GSD under the PIP Framework, most of these would be identified only retrospectively, which adds an additional layer of complexity.

5. This document implies that the use of IVPP GSD outside the GISRS is linked to ‘obligations and/or expectations’ under the present PIP FW, which it is not. Given the complexity in defining and identifying IVPP GSD, which would need to encompass all PIP BM, attempting to impose conditions on the use of IVPP GSD, particularly in the absence of any restrictions to their use, not only appears impractical but would be contrary to the spirit of the PIP FW and may undermine its future effectiveness.

Footnotes:

1 It is noteworthy that the draft document for discussion recently published by the expert Technical Working Group (TWG), entitled “Draft Optimal Characteristics of an Influenza Genetic Sequence Data Sharing System under the PIP Framework” http://www.who.int/influenza/pip/advisory_group/draft_twg_doc.pdf) more accurately reflects the distinctive features of different databases and their GSD sharing mechanisms. Content in that document would appear to supersede certain content in this ‘Options’ document.

2 http://www.ddbj.nig.ac.jp/copyright-e.html
Options to monitor the use of genetic sequence data from influenza viruses with human pandemic potential (IVPP GSD) in end-products

Text in red italics is as in the Draft Paper

Page 1

I) Background

Under the PIP Framework, IVPPs are part of a broader set of materials called ‘PIP Biological Materials’ or ‘PIP BM’, which include human clinical specimens, influenza virus isolates, extracted RNA, cDNA, and influenza candidate vaccine viruses developed from IVPPs by GISRS laboratories.

… the PIP Framework requires that GISRS laboratories submit IVPP genetic sequence data (“GSD”) to “GISAID and GenBank or similar databases in a timely manner”

Since IVPP represent only part of PIP BM, ‘IVPP GSD’ should be defined clearly, in particular as regards the inclusion of sequences from animal viruses that are selected for production of candidate vaccine viruses (CVVs).

If included, how would this be handled retrospectively, as regards obligations and/or expectations under the Framework?

The PIP FW does not make such a statement “requires.” The PIP FW does however state clearly in the Core terms of reference for WHO Collaborating Centers for Influenza and WHO H5 Reference Labs that they:

… upload available … gene sequences of … viruses with pandemic potential to a publicly accessible database” (Annex 5, B5).

The PIP-FW Core terms of reference are clear (not ambiguous) and as the term ‘Core’ defines not just merely a guiding principle for the development of another.

MS have clearly understood that they have a choice as to which publicly accessible database they choose to upload IVPP GSD. The PIP Framework clearly distinguishes the choices by naming specific examples, i.e. Public Domain e.g. GenBank and public access e.g. GISAID to emphasize this distinction.
**Page 2**

**1) Background continued**

The matter has gained importance given the recent development of synthetic biology technologies. In addition to recent developments in synthetic biology technologies, the ease and speed with which sequence data can now be made available (and the amount of data) has also increased the importance of resolving the handling (and sharing) of IVPP GSD as part of the PIP FW.

These advances also have biosecurity implications (which are for example addressed in the modus operandi of GISAID and its database).

... the broader implications of sharing and using IVPP GSD, notably with respect to benefit sharing under the PIP Framework.

As stated above and under section 5.2.4, handling of IVPP GSD as part of the PIP FW has yet to be resolved.
I) Background continued

“The objective of benefit-sharing may be met by monitoring use of GSD and/or tracing GSD or by other mechanisms related to influenza-related products.

While monitoring and tracing the use of GSD is limited by the medium used to share it, technical mechanisms to trace or monitor downloading of GSD from databases may be implemented.

GSD of PIP biological material can also be generated by non-GISRS laboratories.

In that case, WHO will likely not know of this, and the sharing of such will be more difficult to monitor.

… the PIP Advisory Group recommended a process to identify “the optimal characteristics of a system for the handling of IVPP GSD under the Framework, including consideration of […] systems to monitor use of IVPP GSD in end-products”.

… the use of IVPP GSD in relation to the PIP Framework.

Since it is acknowledged that contributions of both upstream and downstream monitoring are important, the upstream options (under section II, 1.1, Pages 4-6) should be summarized more clearly and comprehensively to allow effective evaluation of the relative merits of different options.

This statement is misleading, as it fails to mention the existence and use by MS of a technical mechanism to monitor and trace the use of GSD that has been provided through GISAID database since 2008.

Would this be defined as PIP BM, if outside GISRS and the IVTM?

Sharing of IVPP GSD via a publicly accessible database such as GISAID is possible and routinely done.

Omits consideration a. which deals with assessment of (existing) “data sharing systems that are best suited to meet the objectives of the Framework”, selecting only consideration b., biasing the subsequent discussion of available options.

Handling of IVPP GSD (outside GISRS) as part of the PIP FW has yet to be agreed (section 5.2.4)
II) Discussion

Upstream options should focus on informing entities and individuals accessing GSD of potential obligations and/or expectations under the PIP Framework. This would give more legal certainty to users of the data and could facilitate identification of users of IVPP GSD for benefit-sharing purposes.

1.1. Upstream Options

... mechanisms to monitor use of IVPP GSD to develop end-products, such as vaccines, antivirals and diagnostics. This would allow WHO to identify entities that have used IVPP GSD to generate benefits and result in more transparency for providers of the data.

“GISAID and GenBank or similar databases”

“public open access database”

GSD can be used in many ways to make products. There can be a one to one relationship between a given sequence and a product, or at the other extreme a product can result from knowledge gained from use of many different genetic sequences, which may be provided by GISRS and non-GISRS labs. The latter will be extremely difficult to monitor and verify. Have there been discussions regarding this scenario?

How would merely “informing entities and individuals accessing GSD … give more legal certainty to users … and facilitate identification of users”, without entering into a licensing agreement (with identified users) as for a “public database with data access agreement”, such as GISAID?

The obligations of PIP BM are also not merely communicated through (a notice) informing users, but through bona fide material transfer agreements, e.g. SMTAs

Upstream options should therefore also consider the ability to monitor use of GSD directly through mechanisms (already) provided, for example, by the GISAID database.

“GISAID and GenBank or similar databases”

“public open access database”

See comment 2, regarding the unnecessary ambiguity of this statement.

Versus a private open access database?
This introduction of this novel definition is at best perplexing, as it is neither used (generally) for a category of database, nor used in the definitions stated in the PIP FW (Sec 5.2.2) which distinguishes between two types of publicly accessible databases “public-domain or public-access databases such as Genbank and GISAID respectively”.
Of particular note: GISAID satisfies all conditions on Open Access under the Berlin Declaration as providing Open Access to its data, and “provide the mechanism for enforcement of proper attribution and responsible use of the published work …”
II) Discussion continued

... databases that allow free access to the data they house, without any conditions on the access, use or distribution of data.

It should be noted that every database has some form of ‘conditions’ i.e. ‘Terms of Use’ that are agreed to by the user when accessing the database (website).

For example, GenBank provides such a notice on the ‘Copyright Status’ advising its users that ‘Information ... on this site is within the public domain’. Submitters releasing information into the Public Domain in the United States relinquish all inherent (e.g. Intellectual Property) rights to the data they submit.

Incidentally, “Open Access” as described in the Berlin Declaration, does not preclude databases that require user identification.

Users are not required to register or agree to any terms or conditions to access or use the database.

While it is acknowledged (below) that “it would be difficult to regulate access to data available in such databases in order to implement benefit-sharing under the Framework”, would the attachment of conditions to the use of data not infringe unrestricted rights to use data obtained through ‘public domain’, for example as defined by US legal definition? On what basis could IVPP GSD be singled out for a special concession in attaching restrictions/obligations - is it not inconsistent with the unrestricted/anonymous use of data that has been placed in the ‘public domain’?

... the main disadvantage of public open access databases is inherent in the fact that they are open access. This means that there is no direct tracking of the downloading, distribution, or use of data

This statement is incorrect, given that ‘open access’ to data (e.g. through GISAID), does allow ‘direct tracking of the downloading, distribution, or use of data’ as part of its license data access agreement.

... difficult to regulate access to data available in such databases in order to implement benefit-sharing under the Framework.

Any ‘publicly accessible database’ that by definition provides open access to its data must merely ensure the identity of its users to uphold its ‘terms or conditions’
II) Discussion continued

... request that public open access databases notify their users about PIP Framework obligations and/or expectations."

This would be consistent with the process that is in place for the sharing of PIP biological materials (“PIP BM”)

The PIP Framework does not (yet) impose obligations or expectations on the use of IVPP GSD shared outside GISRS.

This would only be consistent if IVPP GSD were equivalent to PIP BM. It is clearly illogical to separate the two.

... a database may include a general statement that the data may be subject to third-party intellectual property rights.

Once data has been (lawfully) submitted into a public domain database (e.g. GenBank) the owner of the data has relinquished any and all (IP) rights to the data without recourse. Therefore any attempt to link conditions to such data would be invalid, and merely posting a notice on a website provides certainly no legal protection. Removing any and all ‘third-party intellectual property rights’ is at the heart of Public Domain.

How would IVPP be distinguished from similar non-IVPP viruses, unless restricted to viruses from human infections, i.e. subject to the PIP FW?

Such notification statements are generally easily accessible, viewable to all users, and not tied to a particular data entry.

DDBJ (a public domain database) clearly excludes this possibility by stating "The INSDC will not attach statements to records that restrict access to the data ...” (http://www.ddbj.nig.ac.jp/copyright-e.html)

... new statements could be added – to inform users of IVPP GSD of PIP Framework obligations and/or expectations and facilitate their identification.

If acceptable, the WHO PIP Secretariat would need to make a clear statement as to whether or not the use of IVPP GSD would ‘give rise to obligations and/or expectations under the PIP Framework’. Public databases with data access agreement and registration requirement (1.1.2 Page 6) could include such a statement as already developed and offered to the PIP Advisory Group by GISAID (in the form of an trackable information Tag)

The use of genetic sequence data of influenza viruses with pandemic potential may give rise to obligations and/or expectations under the ‘Pandemic Influenza Preparedness Framework for the sharing of influenza viruses and access to vaccines and other benefits” (the PIP Framework) adopted by the World Health Assembly in May 2011.
II) Discussion continued

...notification statements are already in use in many databases provides a useful precedent for implementing this solution. Since public domain databases by default do not conduct positive identification of its users that access GSD, merely posting 'notification statements' somewhere on the website is not helpful, as it would fail to produce any verifiable and auditable acknowledgement of the 'notification' by users accessing IVPP GSD, one that could also be audited.

Many users are still unfamiliar with the PIP FW and those who are will not be thinking about the FW when they are doing their research. Unless the user is specifically reminded that use of IVPP GSD may have some requirements when actually accessing a given sequence they will not be aware. In addition, access to such GSD can never be audited.

GISAID, on the other hand, has a mechanism that not only provides the ability to tag each IVPP GSD, advising users of the special significance of these data, but also gives member states and their authorities the ability to audit the use of the IVPP GSD they submit to GISAID.
II) Discussion continued

The term “public database with data access agreement and registration requirement” refers to databases that are free, publicly accessible and require users to register and agree to certain terms or conditions in order to access and use the database. Terms of use may include a requirement that users acknowledge data contributors, including the laboratory where the original virus sample was obtained as well as the laboratory that generated and submitted the sequence data to the database. Such databases also generally limit further distribution of GSD to other users that have also agreed to the terms of access and use.

Similarly to the solution proposed above for public open access databases, databases with a data access agreement could be asked to include a statement on the PIP Framework on their website or as part of their terms of use.

... identifying the IVPP GSD with accession numbers would greatly facilitate monitoring its use.

This paragraph provides too little information on the distinctive characteristics and advantages of these databases, with respect to linking the sharing of benefits to the use of IVPP GSD.

Although omitted from this section (1.1.2), GISAID also provides the facility to monitor access to and downloading of data, as acknowledged in the report of the TEWG (Oct 2014). As acknowledged on page 7 of this document, monitoring the use of GSD in end products will capture a more limited number of users than monitoring access to IVPP GSD. (“compared to options to monitor access to IVPP GSD, monitoring the use of GSD in end-products will capture a more limited number of users”)

The potential solutions are quite different. Databases with data access agreements can impose conditions, such as obligations/expectations of the PIP FW, for sharing IVPP GSD and may also afford the capacity to monitor and trace the use of GSD (of which there is little mention in this document)

All publicly accessible databases already assign unique ‘accession numbers’ to GSD, since data used in a manuscript submitted for publication would not be (peer) reviewable, resulting in the manuscript to be usually rejected.
II) Discussion continued

1.2. Downstream Options
Downstream options include methods for monitoring the use of IVPP GSD after it has been shared and used to research and develop end-products …

Monitoring the use of IVPP GSD in end-products relies on users clearly and consistently identifying the data in publications, patent applications, or regulatory approval files, with a unique identifier, such as the name of the virus or a database accession number.

Would retrospective identification of obligations of users of Public Domain data not infringe on their rights to unrestricted use of those data?

Would it be not more practical to monitor all GSD rather than specifically monitor IVPP GSD?

It would be more appropriate to always include a database accession number of all GSD.
<table>
<thead>
<tr>
<th>Accession Numbers</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequences can have more than one accession number.</td>
<td>Multiple accession numbers for the same gene segment may arise due to sequences provided by different submitters.</td>
</tr>
<tr>
<td>A potential solution would be to consider technologies to be “end-products”…</td>
<td>What is included under ‘technologies' would need to be further defined, and whether this would apply to e.g. academic researcher who are not developing an end-product and are not obligated under the PIP Framework.</td>
</tr>
</tbody>
</table>
Material and Methods

Ideally, in order to achieve optimal benefit sharing under the PIP Framework, patent applications should include source identification using accession numbers. While it may be difficult to legally mandate this, the practice could be included as part of a best practice guidance for meeting reproducibility requirements.

While this would clearly be best practice, GISAID already requires (acknowledgement of) ‘source identification using accession numbers’.
Material and Methods continued

The source of the Secondary Identifying Number for IVPP GSD could be its accession number. Why not include accession numbers for all GSD as for publication?
Material and Methods continued

This [product] has been developed using nucleotide sequences of influenza viruses with human pandemic potential submitted by [name of laboratory], a member of the Global Influenza Surveillance and Response System (GISRS).

Might this statement not encourage use of IVPP GSD from non-GISRS labs, and undermine all efforts and objectives of the PIP FW?

III) Conclusion

First, identifying uses of IVPP GSD to generate end-products will likely require a combination of both upstream and downstream options. This could include a notice to database users informing them of “PIP Framework Obligations and/or Expectations”.

To encourage this, WHO could issue guidance to users of IVPP GSD specifying that all uses of IVPP GSD to generate an end-product should be acknowledged using accession numbers.

As commented above, the PIP FW does not impose obligations or expectations on the use of IVPP GSD shared outside GISRS, in contrast to PIP BM.

GISAID already requires this condition through its enforceable Terms of Use.
**Conclusion continued**

Lastly, although the objective of this paper is to identify options to monitor the use of IVPP GSD, an alternative could be to monitor uses of IVPPs in all forms (GSD and physical materials).

... instead of tracing accession numbers, WHO could search for mentions of the name of the virus (e.g. A/Anhui/01/2013), which is a unique identifier.

It is of course logical that GSD and the related physical material could be combined as a single entity under the heading PIP BM (or some alternative common designation).

The ability to trace accession numbers yields significantly more information on how IVPP GSD was obtained, and specifically under what terms and conditions it was obtained, e.g. through public domain or through a license issued by a database with data access agreement. Merely using the ‘name of the virus’ would reduce the transparency and hamper the ability to effectively implement certain key objectives of the PIP FW.

The ‘name of the virus’ as an identifier of a virus that would fall under the category of PIP BM, within the context of the PIP FW, which has yet to define how IVPP GSD should be handled.