

1 **WHO’s code of conduct for open and timely sharing of pathogen genetic sequence data**
2 **during outbreaks of infectious disease**

3 According to WHO’s 2016 guidance on managing ethical issues in outbreaks¹, “rapid data sharing is
4 critical during an unfolding health emergency. The ethically appropriate and rapid sharing of data
5 can help identify etiological factors, predict disease spread, evaluate existing and novel treatments,
6 symptomatic care and preventive measures, and guide the deployment of limited resources.”

7 Pathogen genetic sequence data (GSD) is an increasingly valuable source of information in
8 understanding and controlling outbreaks of infectious disease as articulated in a 2017 meeting
9 report of the WHO R&D Blueprint². The extensive applications include, for example, identifying the
10 cause of the outbreak, even if by a previously unknown pathogen, better understanding of
11 transmission, validating diagnostics, and developing therapeutics and preventives. With the advent
12 of next generation sequencing, the depth/extent of available information will expand further.

13 A key concern in recent outbreaks has been variable timelines between the start of an outbreak and
14 the public availability of the first and subsequent genetic sequences. While the first sequences are
15 particularly important to identify and confirm the etiological agent, the ongoing public disclosure of
16 GSD as the outbreak develops allows for monitoring of many elements of the response³.

17 WHO strongly supports public access to sequence data to inform public health and research
18 decision-making during outbreaks, the equitable sharing of benefits derived from the use of such
19 data, and the legitimate interests of data providers. WHO has consulted with many stakeholders and
20 institutions working in the pathogen sequencing arena, including those who have been involved in
21 applications to recent outbreaks. Based on these consultations, and on lessons learned from recent
22 outbreaks as part of the data sharing workstream of the WHO R&D Blueprint, WHO is proposing
23 elements of a code of conduct for GSD sharing in infectious disease outbreaks.

24 Through this code of conduct WHO seeks to contribute towards an enabling environment for sharing
25 of pathogen GSD where equitable benefit sharing and the needs of data providers around the world
26 can be addressed so that rapid international sharing of sequence data can occur consistently, in
27 accord with IHR 2005⁴, allowing public health authorities, product developers and researchers to
28 collaborate more effectively from a position of mutual trust with respect for submitters’ rights to the
29 information they provide.

30 **Recognizing pathogen sequencing as a priority during outbreaks**

31 In relation to disease outbreaks, there should be a commitment by all stakeholders and institutions
32 involved to make publicly available full data on pathogen genomes in order to quickly generate
33 open, interpretable, and actionable information. This should be at no additional cost to the
34 country(s) experiencing the outbreak.

35

¹ <http://www.who.int/ethics/publications/infectious-disease-outbreaks/en/>

² <http://www.who.int/blueprint/meetings-events/meeting-report-pathogen-genetic-sequence-data-sharing.pdf>

³ See International Health Regulations (2005), Article 6. <http://www.who.int/ihr/9789241596664/en/>

⁴ <http://www.who.int/ihr/publications/9789241580496/en/>

36 **Timeliness is critical**

37 Wherever feasible, sequencing should occur as close to the point of sample collection as possible, to
38 empower local capacity, avoid delays, and reduce risk of sample degradation. When the intensity of
39 an outbreak exceeds local capacity, samples should be shared with groups/institutions with
40 demonstrated capacity to generate and share data quickly. The timeframe for data generation and
41 release should not exceed 21 days from sample receipt, although even greater speed is highly
42 desirable in the context of rapidly evolving outbreaks.

43 **Sample export where local capacity is insufficient and cannot be established in a timely manner**

44 Sample export in accordance with national laws is still in the local and global public health interest
45 for certain applications where local capacity does not suffice and cannot be established quickly
46 enough. In such cases a material transfer agreement (MTA) should be used protecting the legitimate
47 interests of the originating country including with respect to issues of ownership, intellectual
48 property and access and benefit sharing. In this connection, the WHO R&D Blueprint has developed
49 a MTA capacity building tool⁵, currently available as a draft for consultation.

50 In cases where sample export occurs, receiving parties should be chosen based on scientific
51 expertise as well as demonstrated commitment to fostering public access to data and results of
52 analyses conducted with the data, respecting terms of ownership, access and benefit sharing,
53 attribution and acknowledgment as well as commitment to local capacity development.
54 International partners should support local capacity development for critical sequence
55 determination and analyses, as a norm associated with working in a developing country.

56 **Early public disclosure of pathogen GSD during each outbreak, followed by journal publication, as
57 agreed under medical journal policies**

58 In each new outbreak or new transmission season GSD should be made publicly available using a
59 public access mechanism, including a brief description of the immediate implications for outbreak
60 control and public health. The uploading of publicly accessible sequence data should occur before
61 journal publication as indicated in WHO's 2015 R&D Blueprint data sharing consultation⁶, and
62 subsequently confirmed in ICMJE policy⁷.

63 In particular the first set of sequences providing crucial information on the pathogen, genotype,
64 lineage, and strain(s) causing the outbreak and which may inform on the origin of the outbreak as
65 well as the choice of diagnostics, therapeutics, and vaccines, should be generated and shared as
66 rapidly as possible. Sharing of corresponding anonymised sample metadata (minimally time and
67 place of collection, ideally with demographic, laboratory, and clinical data) is essential to enhance
68 the interpretation and the value of genomic data.

69 Representatives from leading biomedical journals at a WHO R&D Blueprint consultation on data
70 sharing during public health emergencies made an unequivocal assertion that public disclosure of
71 information of relevance to public health emergencies should not be delayed by publication

⁵ <http://apps.who.int/blueprint/mta-tool/>

⁶ <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001935>

⁷ <http://www.icmje.org/icmje-recommendations.pdf> page 9

72 timelines and that early disclosure should not and will not prejudice journal publication of full
73 scientific reports⁸. WHO advocates that all medical journals urgently update their policies if
74 necessary so that they actively support pre-publication sharing of pathogen GSD related to
75 outbreaks.

76 **Disclaimers preventing publication without coordination with the authors of the original sequence**
77 **disclosure**

78 A common concern is that pre-publication dissemination might allow other scientists to develop
79 publications using such sequence data without involvement of or credit to those who disclosed the
80 original data. One possibility to address this concern is that the providers of the sequence data add a
81 disclaimer notice appended to any disclosure of sequences. A possible template for such a disclaimer
82 is below:

83

DISCLAIMER TEXT; Institutions A & B (from country affected by outbreak) and C & D (international partners) believe in the early sharing of data to support the research and public health response to the outbreak of [name of pathogen or disease]. These data may be downloaded, analysed and used for these purposes. We have released these sequences prior to publication. Our analyses of these sequences is ongoing and a publication is in preparation. It is not permitted to use the sequences for publication in any type of communication with the general public that describes data generated with the use of the sequences. If you intend to so, you must obtain our permission and coordinate with us.

84 This approach has been effectively used in several outbreaks.

85 Publication based on publicly available sequence data prior to publication of the results by the
86 originators without collaboration with those who originally generated the sequences will be
87 considered contrary to this code of conduct.

88 **Access to analysis capacity**

89 In some settings the main bottleneck to provision of interpretable results from sequencing efforts is
90 the lack of bioinformatics expertise. In some cases sequencing databases can run to terabytes of
91 data, such that raw data transfer and analysis can become a more limiting factor than sequencing
92 itself. An international collaborative model is desirable for development of easy-to-operate, open
93 source algorithms to allow for on-site or cloud-based analysis and interpretation of locally generated
94 sequence data. However until such time as this model is functioning, a network of partners
95 supporting analysis of locally generated data is needed. WHO can assist in establishing such a
96 network.

97

⁸ <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001935>

98 **Incentives**

99 Public and philanthropic funders of sequencing initiatives should incentivise groups following the
100 above norms on sequence data access as appropriate.

101 **Models for sharing pathogen sequence data**

102 Alternative options for publicly accessible pathogen sequence data sharing platforms should be
103 available to serve the needs and preferences in different situations.

104 Where data providers are not concerned about retention of ownership of the data, databases
105 without data access agreements (such as GenBank, ENA and DDBJ) are the default option for
106 sharing.

107 For situations where data providers seek retention of ownership of their data, alternative models
108 with data access agreements (such as GISAID⁹ in influenza) have been used to facilitate rapid sharing
109 of GSD.

110 The following elements could be used in such models, with provisions on ownership and access and
111 benefit sharing:

112 1) A database for hosting pathogen sequence data that meets best practice standards for
113 administration and security

114 2) A database access agreement in the form of a simple click-through interface that allows
115 'providers' to share data and allows any genuine individual receiver party ('user') to access the data
116 in alignment with relevant collaboration principles as outlined in the terms and conditions of the
117 data access agreement, including provisions on ownership, intellectual property and access and
118 benefit sharing. These conditions should include the following elements:

- 119
- 120 • allow for ownership of the sequence data to reside with the provider of the data uploaded
to the platform;
 - 121 • anyone using data accessed from the platform is required to acknowledge/credit/potentially
122 co-author with the providers as appropriate;
 - 123 • (secondary) users should, as appropriate, propose and seek to collaborate with the data
124 providers, including joint analysis of data;
 - 125 • appropriate intellectual property management

126 3) A governance mechanism which includes appropriate handling of any conflict of interests and
127 allows for engagement and trust between all parties, including low and middle income countries.
128 This governance mechanism could include the possibility of restricting downloads from the database
129 to those that do not follow required best practices.

130 4) A dispute resolution mechanism that can mediate where disagreements arise.

131 It is proposed that WHO explore with partners how to establish such mechanisms for sharing GSD
132 for pathogens, and engage in discussions as to where an associated database might be hosted.

⁹ Global Initiative on Sharing All Influenza Data

133

134 **Monitoring and evaluation of this approach**

135 The extent to which open, timely, pre-publication sharing of sequence and metadata occurs during
136 outbreaks would be monitored. At the same time, the extent to which principles of equitable benefit
137 sharing and recognition of data providers' legitimate interests are honoured should also be
138 evaluated. The success of this approach will depend on its effective functioning as a balanced,
139 credible and equitable system that advances global public health objectives and addresses the
140 legitimate interests of both users and providers of data.

141

DRAFT

142 Table 1:

143 **Uses of pathogen sequence data in outbreaks:**

- 144 1. Characterizing an outbreak pathogen directly or through the identification of the
- 145 pathogen's common ancestor if it is a novel pathogen;
- 146 2. Determining the molecular epidemiology of the pathogen, including its diversity,
- 147 virulence, transmission routes and timelines, and possible mutation;
- 148 3. Identifying hotspots of transmission and helping to pinpoint outbreaks to map
- 149 where emergency response measures are needed most,
- 150 4. Providing clarity on contentious aspects of a given outbreak by specifying the route
- 151 of transmission within humans and between humans and potential zoonotic
- 152 reservoirs. This can help target response measures effectively to the most affected
- 153 or at risk areas, and trace the origin of an outbreak more accurately;
- 154 5. Validating and improving molecular diagnostic tool development (i.e. PCR-based
- 155 assays) based on up to date, relevant genomic data of given pathogen, especially
- 156 early in the outbreak;
- 157 6. Underpin R&D to better understand the immunopathology of the disease, and to
- 158 identify and/or develop useful interventions (therapeutics and vaccines) for
- 159 immediate or future outbreak control and prevention;
- 160 7. Identify sources and sinks of infections, distinguish locally-transmitted versus
- 161 imported cases;
- 162 8. Improve epidemiological models predicting disease transmission.

163 **Longer term benefits:**

- 164 9. Improving disease surveillance in human and animal hosts. Genome-based
- 165 surveillance through portable technology can improve diagnostic capabilities in
- 166 resource-poor (e.g. field) settings, and play a role in detecting outbreaks earlier;
- 167 10. Facilitating post emergency research and improving the evidence base of "lessons
- 168 learned." This in turn aids preparedness for the next outbreak, wherever it takes
- 169 place. Epidemics are not bound by national borders, and sharing GSD from one
- 170 country allows collaborative exchanges of ideas, products, and expertise which may
- 171 be beneficial for a similar epidemic in a different country;
- 172 11. Monitoring resistance to therapeutics such as antivirals;
- 173 12. Allowing the self-correcting nature of science to progress. Initial interpretations of
- 174 data can be mistaken, and data availability allows replication and peer review. For
- 175 translational research that leads to public health decisions, a wrong conclusion
- 176 without prompt correction can slow progress significantly.

177

178

179

180

181