



# WHO guiding principles for pathogen genome data sharing

## WHO Guiding principles for pathogen genome data sharing

ISBN 978-92-4-006174-3 (electronic version)

ISBN 978-92-4-006175-0 (print version)

© World Health Organization 2022

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (<http://www.wipo.int/amc/en/mediation/rules/>).

**Suggested citation.** WHO Guiding principles for pathogen genome data sharing. Geneva: World Health Organization; 2022. Licence: [CC BY-NC-SA 3.0 IGO](https://creativecommons.org/licenses/by-nc-sa/3.0/igo).

**Cataloguing-in-Publication (CIP) data.** CIP data are available at <http://apps.who.int/iris>.

**Sales, rights and licensing.** To purchase WHO publications, see <http://apps.who.int/bookorders>. To submit requests for commercial use and queries on rights and licensing, see <https://www.who.int/copyright>.

**Third-party materials.** If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

**General disclaimers.** The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

---

# FOREWORD

---



**Dr Tedros Adhanom Ghebreyesus**  
Director-General  
World Health Organization

One of the defining features of the COVID-19 pandemic is the lack of solidarity and sharing that has made the global response to the most severe health crisis in a century fractured and uncoordinated.

The world needs timely, high quality and geographically representative sharing of pathogen genome data in as close to real time as possible. When pathogen genome data is shared nationally and internationally, it helps to prevent, detect, and respond to epidemics and pandemics. Regular collection and sharing of pathogen genome data is also crucial for endemic diseases, especially for pathogens that are resistant to antimicrobials and require regularly updated policies. Genomic surveillance is critical for early warning of new epidemics, to monitor the evolution of infectious disease agents, and develop diagnostics, medicines and vaccines. This technology has been crucial in our response to the COVID-19 pandemic, from identifying a novel coronavirus to developing the first diagnostic tests and vaccines, to tracking and identifying new variants.

Capacities for pathogen genomics around the world are advancing rapidly. The speed, scale and affordability of sequencing are all increasing at astonishing rates. While these advances are welcome, inequities in access to new technologies leave blind spots in global surveillance that put us all at risk. A commitment to equity and support for capacity development must therefore be the overriding principle of the international system for sharing of pathogen data.

Even when countries have the technology and capacity to perform high quality genomic sequencing, the lack of global standards or rules for sharing genomic data can disincentivize them from sharing those data, out of concern for the potentially negative effects of reporting the emergence of a new and dangerous pathogen on trade and tourism. During the COVID-19 pandemic, some countries that shared high quality data on new variants were punished with travel restrictions, instead of being applauded for their act of solidarity.

The publication of the WHO Guiding Principles for Pathogen Genome Data Sharing is therefore very timely, providing practical assistance to researchers, epidemiologists, and public health officials. It charts a path for timely sharing of pathogen genome data, while acknowledging the legitimate concerns and needs of scientists around the world that are the originators of this data.

As the world recovers from the COVID-19 pandemic and begins to build a stronger global architecture for health emergency preparedness and response, these guiding principles are a much-needed tool for constructing a more equitable and transparent global system that keeps us all safer.

A handwritten signature in blue ink, which appears to read "Tedros Adhanom". The signature is written in a cursive style.

# ACKNOWLEDGEMENTS

---

These principles for the sharing of pathogen genome data have been developed building upon experience with recent outbreaks of infectious disease and after extensive consultation with experts. The most recent consultation was held in April 2022 at the Rockefeller Foundation Bellagio Conference Center (see Annex for List of participants). In addition, WHO would like to acknowledge the other stakeholders who took time to provide comments. This includes representatives from national public health agencies, from international organizations, from laboratory networks, from international associations (including private sector), research institutions and research funding organizations, and WHO staff from headquarters and regional offices.

# INTRODUCTION

---



WHO encourages the sharing of pathogen genome data to protect global public health. Sharing of pathogen genome data is critical for preventing, detecting, and responding to epidemics and pandemics at national and international levels, and is in the interest of all Member States. The regular collection and sharing of such data are also important for monitoring and responding to endemic diseases and for tracking antimicrobial resistance to inform policy decisions.

Practices and policies for sharing pathogen genome data must be ethical, equitable, efficient, and effective. After wide consultation, WHO has developed these foundational principles, which focus on public health uses, as well as urgent immediate research priorities.

## Scope and purpose

These principles apply to the rapid, public sharing of pathogen genome data, along with appropriate metadata required for interpretation (excluding any sensitive clinical data). They do not apply to non-pathogen genome data. The principles support the enhanced, timely sharing of quality

data, both within countries and internationally, and will contribute to capacity development and equitable access to benefits that arise from the use of these data.

Although the principles apply to the genomes of organisms pathogenic to humans, they may also have some applicability to animal pathogens with the potential to spillover into humans.

These principles are aligned with the 2021 “WHO Global Genomics Surveillance Strategy for pathogens with pandemic and epidemic potential”, specifically Objective 3: “Enhance data sharing and utility for streamlined local to global public health decision-making and action<sup>1</sup>”.

## Audience

The target audience for this document is technical staff at public health agencies and researchers involved in the prevention and control of endemic, epidemic, and pandemic diseases. The principles can be adopted by those who generate data, as well as those who operate or use data-sharing platforms.

---

<sup>1</sup> <https://apps.who.int/iris/handle/10665/352580>

# GUIDING PRINCIPLES

---



## 1. Capacity development

Member States increasingly prefer local analysis of data as the basis for their decision-making. The generation of high-quality pathogen genome sequencing data that can be shared quickly and effectively in a global system requires capacity and infrastructure.

Any pathogen genomics initiative should therefore contribute to capacity development that can establish and sustain data generation, processing, and submission, as well as reliable curation and annotation. In this manner, all countries will be able to analyze data.

Additional resources may be needed to improve infrastructure. For example, laboratories may require upgrades to electricity, regulated ambient

conditions, and internet access; additional equipment and supplies may be needed; personnel training may require strengthening and promotion<sup>2</sup>; additional computational capacity (e.g., servers or cloud-based platforms) and informatics may be needed to sustain high-throughput sequence production with appropriate quality controls.

## 2. Collaboration and cooperation

A global system of pathogen genome sharing should promote collaboration and cooperation between the laboratories and institutions that submit data to it and the scientists who analyze the data and report their analyses. Further, collaborations should seek to develop local analysis capacity in countries, including local infrastructure, informatics, and data management functions.

---

<sup>2</sup> Examples of other WHO documents that provide more details on operational considerations:

[“GLASS whole-genome sequencing for surveillance of antimicrobial resistance”](#)

[“Operational considerations to expedite genomic sequencing component of GISRS surveillance of SARS-CoV-2”](#)

[“Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health”](#)

### 3. High-quality, reproducible data

To ensure that data analyses are optimal for decision-making, close attention should be paid to the quality of the data and supporting metadata. This requires support for data curation and quality assurance. In emergencies, however, there is a trade-off between the time needed to attain very high-quality data and the ability to share data rapidly. In some cases, it may be appropriate to share lower quality data (clearly marked as “preliminary/not fully quality controlled” or similar), with the understanding that preliminary data need to be easily identifiable. The development, implementation, and maintenance of data standards should be prioritized for robust data sharing. Where possible, external quality assessments should be performed as part of capacity building efforts.

Raw sequencing reads linked to final consensus sequences should be submitted to facilitate peer review wherever possible. This is particularly important during the initial phase of outbreaks and when the emergence of new variants or recombinants is observed. Human genomic data that may inadvertently be included in raw reads should be removed before submission to or release of the sequence on public databases.

### 4. Global and regional representativeness

The sharing of pathogen genome sequences by as many countries as possible is critical for global analyses that can identify the emergence of new infectious threats.

At the global level, the first few sequences submitted by countries that have not previously shared can be extremely informative. Timeliness and geographical and temporal representation relative to disease epidemiology govern the public health value of pathogen genomic data. Therefore, a global system of pathogen genome data sharing must address the needs of all countries.

### 5. Timeliness

During epidemics and pandemics, generation of pathogen genome data and sharing of such data should be as timely as possible in order to make data available for analyses. The sharing of pathogen genome data should not be delayed because a data submitter wishes to prepare a scientific manuscript for publication. But neither should data submitters be obliged to waive their rights to the data they share before publication. The source of all data should be acknowledged in all published results. Scientific journals should encourage, and not prevent, the release of pathogen sequencing data before publication.

### 6. Acknowledgement and intellectual credit

All contributions, including those of submitting laboratories, and where appropriate, laboratories from which clinical samples or pathogen isolates have originated, should be appropriately acknowledged in presentations and publications. Where pathogen genome data from global platforms are used in scientific research projects, scientists from the submitting and originating laboratories should be invited to participate in that research. Moreover, the data users should make good-faith efforts to engage the submitting and originating laboratories in the preparation of manuscripts for presentation and publication. Authorship should follow the guidelines of the International Committee of Medical Journal Editors.

### 7. Equitable access to health technologies as a benefit

Equity in access to health technologies is an important potential benefit of the sharing of pathogen genome data, but timely sharing does not

**“Scientific journals should encourage, and not prevent, the release of pathogen sequencing data before publication”**

necessarily lead to equitable access to derivative benefits. Benefits may arise from individual data submissions as well as from entire datasets (i.e., the entire dataset is required for the statistical analysis of variants or the development of artificial intelligence-powered tools). Data-sharing policies and practices should contribute to equitable access to health technologies wherever possible. Although pathogen genome sequence sharing platforms may not be able to directly ensure access to medical interventions, such platforms should set expectations that all users of the data apply equity and fairness considerations in their use of the data for developing health technologies. This should include equity in access to sequencing and computing and analysis technologies involved with generating, curating, uploading, downloading, and analyzing data, as well as other aspects of equity discussed in this document.<sup>3</sup>

**“Whenever possible, pathogen genome data should be made available in a timely manner on publicly accessible platforms that provide either anonymous or verified user access”**

## 8. As open as possible and as closed as necessary

Whenever possible, pathogen genome data should be made available in a timely manner on publicly accessible platforms that provide either anonymous or verified user access.

Unrestricted open access models for sharing should be used whenever data submitters do not wish rights to be reserved. These allow anonymous access, unhindered aggregation and use of data for research and commercial applications.

Where submitters opt to retain certain protections related to pathogen genome data, platforms that preserve data generators’ rights should be used. In this case, any user seeking access should be granted such access under agreed terms. Access should be free of charge for users.

Both models with publicly accessible data allow for global analyses and development of diagnostics, medicines and vaccines. Both types of platforms require sustainable financing mechanisms.

## 9. Interoperability and relevance for national, regional and global decision-makers

Ideally, pathogen genome sequence data should be submitted in a form that permits aggregation, to avoid multiple partitioned datasets. Achieving this requires compatible access and usage terms and technical interoperability between platforms, while the rights of data providers are respected. Institutions within the global system of pathogen genome data sharing that operate different sharing platforms should collaborate on identifier systems and data reporting standards. The global system of pathogen genome data sharing should allow for large-scale standardized analyses that follow good practices developed in a rigorous manner by the global scientific community. This will permit answers to critical public health questions such as epidemiological characteristics, transmission dynamics, emergence of variants, drug or vaccine resistance and the effectiveness of diagnostics.

Where the retention of data ownership and benefits such as acknowledgement, attribution, and supporting scientific collaboration are guaranteed to data submitters, terms and conditions should include enforceable compliance (see 13 overleaf).

---

<sup>3</sup> Global and regional efforts are ongoing to advance affordable access to new products (diagnostics, therapeutics, vaccines) so that these benefits flow to countries in medical need including those from which data are generated.



## 10. Trustworthiness and ease of use

Global pathogen genome data sharing practice should prioritize developing and sustaining trust among providers, platforms, and users of the data. Establishing who controls data sharing and where data are stored is important for achieving trust. Where all parties have confidence in the global system of pathogen genome data sharing, consider it fair,<sup>4</sup> and support the underpinning principles, data sharing will be enhanced. Conversely, without such trust, it is unlikely that quality data will be submitted rapidly before publication.

Platforms should prioritize ease of data submission, while supporting data curation that allows for sound analyses.

## 11. Transparency

Terms and conditions of databases in the global system of pathogen genome data sharing should be publicly available. Governance procedures, for example, in the constitution of boards and committees and their terms of reference, should be publicly available. The governance structure should strive to ensure efficient and effective administration of the data sharing platforms while aligning the functioning of platforms with the principles outlined here.<sup>5</sup>

## 12. Consistency with applicable law and ethical regulations

Pathogen genome data sharing platforms should be established and operated in a manner consistent with applicable national and international laws, regulations, rules, and standards, and ethical regulations, norms, and standards.

## 13. Compliance and enforcement

Mechanisms to ensure compliance with principles are needed to make them credible. Data sharing platforms whose terms and conditions provide specific protections to data submitters, should include mechanisms and resources to ensure compliance. Where data sharing mechanisms permit, there may also be a need to raise awareness of instances and sanctions for breaches of the rules of data access as a means of supporting trustworthiness.

**“Pathogen genome data sharing platforms should be established and operated in a manner consistent with applicable national and international laws, regulations, rules, and standards, and ethical regulations, norms, and standards”**

---

<sup>4</sup> Fair here is not the acronym FAIR. Discussions held by WHO have highlighted the importance of addressing equity in addition to aligning with FAIR data sharing principles (Findable, Accessible, Interoperable, Reusable) leading to the suggestion that FAIR should be replaced with FAIR+E. Further work is needed to elaborate how data sharing can appropriately support equity.

<sup>5</sup> Not all principles will apply to all platforms. For example, where pathogen genome data sharing is intended to be unrestricted, with no rights reserved, sections on requirements on retaining rights and compliance and enforcement are less relevant.

# ANNEX

## List of participants for meeting held at the Bellagio Center, April 25-28, 2022

Name	Organization/Institution	Country
------	--------------------------	---------

### Representative of countries/regions

Budi Gunadi Sadikin	Ministry of Health	Indonesia
Rajesh S Gokhale	Secretary, Department of Biotechnology	India
Li Mingkun	Beijing Institute of Genomics, Chinese Academy of Sciences	China
Raymond Lin Tzer Pin	National Public Health Laboratory, Ministry of Health	Singapore
Benjamin Howden	Public Health Laboratory, The University of Melbourne	Australia
Steve Sherry	National Center for Biotechnology Information, National Institutes of Health	USA
Marilda Siqueira	Oswaldo Cruz Foundation, FIOCRUZ, Ministry of Health, Brazil	Brazil
Josefina Campos	Administración Nacional de Laboratorios e Institutos de Salud, Buenos Aires	Argentina
Nawal Al Kaabi	UAE Genome Program	United Arab Emirates
Nada Melhem	American University in Beirut	Lebanon
Meera Chand	UK HSA	United Kingdom
Torsten Semmler	Robert Koch Institute	Germany
Marie-Paule Kieny	INSERM	France
Erik Alm	European Centre for Disease Prevention and Control (ECDC)	Sweden
Sofonias Kifle Tessema	Africa CDC	Ethiopia
Amadou Sall	Institut Pasteur, Dakar, Senegal	Senegal
Tulio D'Oliveira	Centre for Epidemic Response & innovation, South Africa	South Africa
Giuseppe Ippolito	Ministry of Health	Italy

### Leader of existing platform

Peter Bogner	GISAID	USA
Rolf Apweiler	European Bioinformatics Institute	United Kingdom

# ANNEX

---

Name	Organization/Institution	Country
------	--------------------------	---------

## Other key stakeholders

Charlotte Germain-Aubrey	Secretariat of the Convention on Biological Diversity	Canada
--------------------------	---	--------

## Facilitator

Suerie Moon	Graduate Institute	Switzerland
-------------	--------------------	-------------

## Academic Stakeholder

Rebecca Katz	Centre for Global Health Science & Security, Georgetown university	USA
--------------	--	-----

## Funder Stakeholders

Naveen Rao	Rockefeller foundation	USA
Rick Bright	Rockefeller foundation	USA
Bruce Gellin	Rockefeller foundation	USA
Tariq Khokhar	Wellcome Trust	United Kingdom
Trevor Mundel	Bill & Melinda Gates Foundation	USA

## WHO staff

Vasee Moorthy	WHO
Oliver Morgan	WHO
Soumya Swaminathan	WHO
Chikwe Ihekweazu	WHO

WHO staff members who contributed to this document include: Vasee Moorthy in a coordinating role, Oliver Morgan, Mark Perkins, Jilian Sacks, Wenqing Zhang, Matteo Zignol, Nazir Ismail, Pascal Ringwald, Anthony Solomon, Carmem L. Pessoa da Silva, Jorge Matheu Alvarez, Silvia Bertagnolio, Mick Mulders, and Ousmane Diop, as well as staff from WHO Regional Offices. Senior leadership was provided by WHO Chief Scientist Soumya Swaminathan and Assistant Director-General Chikwe Ihekweazu. Direct financial and logistical support was provided by the Rockefeller Foundation; WHO retained full control of the agenda and list of participants.

9789240061743



9 789240 061743