World Health Organization

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)

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.

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.

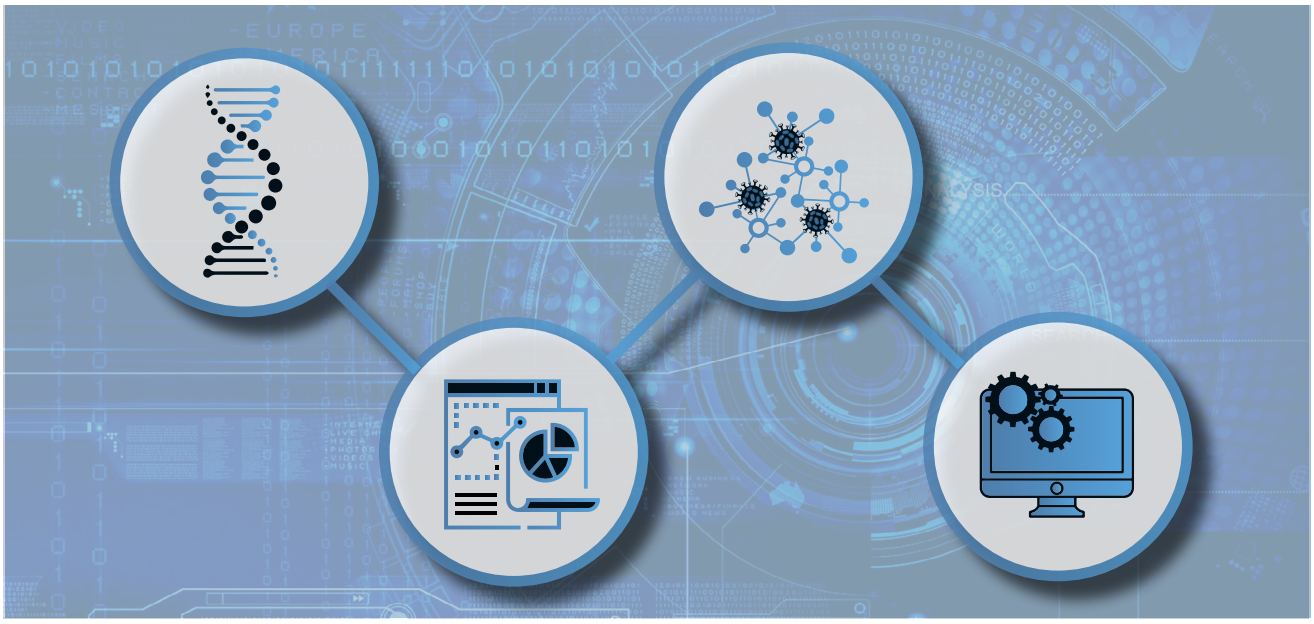
These principles are aligned with the 2021 "WHO Global Genomics Surveillance Strategy for pathogens with pandemic and epidemic potential", specifically Objective 3: "Enhance data sharing and utility for streamlined local to global public health decision-making and action[1]".

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.

[1] 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[2]; 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.

[2] 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

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