Postprint: Research on Data Sharing Mechanisms in Novel Coronavirus Public Health Emergencies

Authors: Cui Yuhong, Wang Sa

Date: 2023-04-01T16:15:58+00:00

Abstract

[Purpose/Significance] Amidst the COVID-19 pandemic, establishing and improving open data platforms and sharing mechanisms to address global public health challenges and advance the development of data sharing capabilities for public health emergencies has become a widespread consensus among governments and the scientific community worldwide.

[Method/Process] First, we examine the data sharing policy frameworks and response measures in previous public health emergencies. Subsequently, through an empirical investigation of data sharing sources including peer-reviewed journals, preprints, data repositories, clinical trial registration platforms, and genomic/structural data centers, we analyze the current state of COVID-19 data sharing. Finally, we discuss the advantages, disadvantages, and obstacles of different data sharing mechanisms in public health emergencies.

[Results/Conclusions] The degree of data sharing during the COVID-19 pandemic has significantly improved, yet has not become normalized. Technical, motivational, economic, political, legal, and ethical factors represent the primary obstacles to data sharing. China should enhance its data sharing capabilities in public health emergencies through strategic planning, infrastructure development, stakeholder engagement, and ethical-legal frameworks.

Full Text

Abstract

Amid the COVID-19 pandemic, establishing and improving open data platforms and sharing mechanisms to address global public health challenges and enhance capacity for data sharing in public health emergencies has become a widespread

consensus among governments and the scientific community. This study first examines data sharing policies and response measures from previous public health emergencies, then analyzes the current state of COVID-19 data sharing through an empirical investigation of peer-reviewed journals, preprints, data repositories, clinical trial registration platforms, and genome/structure data centers. Finally, it explores the advantages, disadvantages, and obstacles of different data sharing mechanisms during public health emergencies. The findings reveal that while COVID-19 data sharing has improved significantly compared to previous outbreaks, it has not yet become normalized. Technical, motivational, economic, political, legal, and ethical factors constitute the main barriers to data sharing. China should strengthen its public health emergency data sharing capacity through strategic planning, infrastructure development, stakeholder coordination, and ethical-legal frameworks.

Keywords: data sharing mechanisms; public health emergency; COVID-19

On January 30, 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a public health emergency of international concern. During such emergencies, data sharing is crucial for developing effective public health emergency management strategies. Data forms the foundation of public health action, and successful data sharing policies can facilitate early disease prevention, timely global communication of clinical trial data and molecular epidemiological information, and appropriate interventions to avoid catastrophic outcomes such as mass illness or death. Recent global outbreaks—including the 2009 H1N1 influenza pandemic, the 2013–2016 West African Ebola epidemic, and the 2015 Zika virus syndrome—have sparked widespread discussion about research data sharing in public health. The Global Research Collaboration for Infectious Disease Preparedness (Glopid-R) assessed data sharing practices for pathogens with pandemic potential by reviewing literature and policies from 12 major pathogens over two decades. To better understand barriers to data sharing during emergencies, the Wellcome Trust and UK Department for International Development commissioned Glopid-R's data sharing working group to develop six case studies highlighting implementation complexities across different outbreak scenarios. According to *Nature*, at least 80 papers were published within the first month of the COVID-19 outbreak, and statistics through May 2020 show exponential growth in COVID-19 research publications. The virus's rapid global spread has underscored challenges facing public health data sharing practices. In response, funding agencies and publishers quickly implemented open access policies for COVID-19 research. This study examines data sharing policies in major public health emergencies and investigates current COVID-19 data sharing practices to explore characteristics and obstacles of different mechanisms, aiming to advance China's capacity for public health emergency data sharing.

1. Data Sharing Norms and Measures in Major Public Health Emergencies

Over the past two decades, emerging and re-emerging viral infections and pandemics have posed significant threats to global public health security. This section examines several major outbreaks that garnered widespread public attention, including the 1997 H7N9 influenza in China and other Asian countries, the 2009 H1N1 pandemic, the 2013–2016 Ebola epidemic, the 2015 Zika virus syndrome, and the COVID-19 outbreak that began in late 2019. WHO, national funding agencies, and top publishers have responded rapidly to these crises, substantially promoting the establishment, improvement, and implementation of global data sharing initiatives, international norms, and journal publication guidelines.

1.1 Global Initiative on Sharing All Influenza Data (GISAID)

In 1997, human infections with avian influenza and other zoonotic influenza viruses—such as H5N1, H7N9, and H9N2 subtypes—raised serious concerns about a potential influenza pandemic. Alarmingly, when the first H5N1 cases were identified in Indonesia, WHO had little access to up-to-date animal sequence data to understand viral evolution. Priority disputes over publication led researchers to withhold raw materials and data on emerging diseases, and even the H5N1 database established with U.S. CDC funding restricted access to countries that contributed sequences. To address these challenges, over 70 prominent scientists published a joint call in *Nature* in 2006 to establish the Global Initiative on Sharing All Influenza Data (GISAID), requiring all data to be released to GenBank and other public archives within six months of submission, with all users adhering to intellectual property and attribution rules.

Traditional public domain archives have proven inadequate for rapid data sharing during pandemics, while GISAID provides an effective and trusted mechanism for sharing both published and unpublished influenza data. In 2008, GISAID launched the EpiFlu database, enabling countries to track new viral spread and evolution globally—a capability that proved vital during the 2009 H1N1 pandemic. In 2013, Chinese scientists were the first to publish H7N9 virus gene sequences, which were used within weeks to develop and test vaccine viruses through synthetic biology experiments, demonstrating acceptance and trust in GISAID's sharing mechanism and proving its feasibility for timely sharing of critical influenza data. In March 2020, the China National GeneBank partnered with GISAID, making its life science data platform the first officially authorized GISAID platform in China.

1.2 International Data Sharing Norms

The massive West African Ebola outbreak (2014–2016) and the Zika virus infection crisis (2015), which caused Guillain-Barré syndrome and microcephaly, significantly accelerated the development of international data sharing norms.

In September 2015, WHO convened a consultation to establish global norms for providing data and research findings as openly and rapidly as possible during public health emergencies. Addressing concerns about pre-publication disclosure of critical information, WHO clarified that journals should encourage or require data sharing before article publication during emergencies, including laboratory results (genomic and immunological data), demographic statistics, animal study results, and clinical participant data.

To support implementation of WHO's global data sharing norms, 31 institutions—including the UK Academy of Medical Sciences, NIH, NSF, Wellcome Trust, Bill & Melinda Gates Foundation, Biotechnology and Biological Sciences Research Council, and major journals such as *BMJ*, *The Lancet*, *Science*, and PLOS—signed a joint statement in February 2016. The statement committed to improving current and future public health emergency data sharing mechanisms to reduce delays in sharing research outcomes, emphasizing implementation pathways for data sharing during emergencies. Journals pledged to provide free access to all Zika-related content, and data or preprints submitted for unrestricted dissemination would not be considered prior publication. Funders were required to establish appropriate mechanisms for researchers to share quality-assured interim and final data as widely and rapidly as possible with public health and research communities and WHO.

The International Committee of Medical Journal Editors (ICMJE) announced in January 2016 that clinical trial reports submitted to ICMJE journals after July 2018 must include data sharing statements, and trials enrolling participants after January 2019 must submit data sharing plans to clinical trial registries. Any changes to these plans must be updated on registration platforms and explained upon manuscript submission.

The "Enhancing Public Health Data Sharing" project, led by the Center for Global Health Security and supported by the Bill & Melinda Gates Foundation, organized a series of roundtables to exchange experiences on creating proper data sharing environments. The project developed the "Guide to Sharing Public Health Surveillance Data and Benefits," targeting global health partners—including public health departments, NGOs, private sector entities, academic institutions, and multilateral organizations—to promote international data sharing norms as a model for sharing data as openly and appropriately as possible.

1.3 Policy Responses Since the COVID-19 Outbreak

Since December 2019, COVID-19's rapid spread has posed a major threat to global health. To address barriers to data access during outbreaks, the *WHO Bulletin* implemented the "COVID-19 Open" data sharing and reporting mechanism. When data is submitted to the *Bulletin*, all COVID-19-related research papers are assigned a digital object identifier and posted online within 24 hours of acceptance for peer review on the "nCoV-2019 Open" data platform. Under Creative Commons licensing, data ownership remains with authors, and the

data can be freely used, distributed, and reproduced in any medium provided the original work is appropriately cited. If a paper passes peer review, the open review results are reported in the final publication. If not, authors may submit to other journals or publish on rapid sharing platforms recommended by WHO, such as ProMED and F1000Research.

The Wellcome Trust, *Nature*, and other publishers, funders, and scientific societies signed a joint statement to ensure rapid sharing of coronavirus-related research data and findings. Key principles include: (1) all outbreak-related peer-reviewed papers must be immediately open access; (2) outbreak-related findings must be shared with WHO immediately upon journal submission; (3) journals may obtain research results through preprint services and other open access platforms before formal publication or peer review; (4) researchers, public health institutions, and WHO should share outbreak-related data, protocols, and standards as quickly and widely as possible; and (5) shared data or preprints will not be considered prior publication by journals.

At the WHO "Global Research and Innovation Forum on COVID-19" held on February 11–12, 2020, data sharing was listed as one of eight priority research agendas. Scientists worldwide agreed that viral resources, clinical samples, and related data should be shared rapidly for public health practice, and that medical products or innovations developed from these resources must also be shared equitably and fairly.

2. Data Sharing in the COVID-19 Public Health Emergency

To understand the current state of data sharing since the COVID-19 outbreak, this study examined five data sharing sources: peer-reviewed journals, preprints, data repositories, clinical trial registries, and genome/structure data centers. Search terms included "coronavirus," "2019-nCoV," or "COVID-19," with the timeframe limited to January 1–May 29, 2020.

2.1 Peer-Reviewed Journals

From the perspective of scientific research integrity, the peer review system effectively ensures the quality and reliability of shared research. Four major publisher platforms—Elsevier, Wiley, Springer, and Taylor & Francis—were examined, along with five specialty journals: *The Lancet*, *Journal of Medical Virology* (JMEDVIROL), *Nature*, *New England Journal of Medicine* (NEJM), and the *Chinese Journal of Tuberculosis and Respiratory Diseases*. After retrieving and manually screening COVID-19-related literature (counting only "articles" and "reviews"), each paper was checked for data sharing statements and supplementary materials . Papers providing underlying data as supplementary materials or with statements on data accessibility were counted as having shared data.

Nearly all COVID-19 journal papers are accessible through open access or free full-text access. As shown in Table 1, approximately 20% of COVID-19 papers on the four major publisher platforms include shared data, though sharing methods and extent vary by platform. Elsevier published the highest number of COVID-19 papers and shared the most data, providing data statements, accessibility information, and supplementary materials in formats including PDF, tables, videos, audio, images, and software code. Data statements either indicate that researchers will determine sharing based on user requests or provide sources and accessibility for data used in papers.

Differences in data sharing among individual journals are more pronounced than among publisher platforms. As shown in Table 2, *The Lancet*, *Nature*, and *NEJM* have significantly higher data sharing rates than publisher platforms and other journals, while the *Chinese Journal of Tuberculosis and Respiratory Diseases* provides no data sharing mechanisms for its COVID-19 papers.

2.2 Preprints

To accelerate widespread dissemination of COVID-19 research, researchers may submit manuscripts to preprint platforms before formal journal acceptance. However, these unreviewed preliminary studies should not guide clinical practice or health behaviors, nor be reported as established facts in news media. A notable example involved a January 31 preprint on bioRxiv by researchers from the Indian Institute of Technology that was widely circulated but later strongly questioned by academics for exaggeration and distortion, ultimately leading to retraction.

Six preprint platforms were examined: arXiv, SSRN, bioRxiv, chemRxiv, medRxiv, and ChinaXiv. After retrieving and manually screening COVID-19 literature, each platform's data sharing functionality was assessed . As shown in Table 3, data sharing practices on preprint platforms correlate significantly with disciplinary focus. Life and health science preprints play the primary role in COVID-19 data sharing, while the two long-established services—arXiv (physics and mathematics) and SSRN (social sciences and humanities)—lack functionality for sharing supplementary documents. SSRN was acquired by Elsevier in 2016 and has partnered with *The Lancet* and *Cell* since 2018, allowing authors submitting to these journals to opt for rapid preprint sharing on SSRN, though SSRN provides no data access.

bioRxiv, chemRxiv, and medRxiv are newer preprint services for life and health sciences. Since the outbreak, ChinaXiv has fully supported rapid publication of domestic and international COVID-19 research, allowing data downloads as attachments. Data sharing proportions on these platforms far exceed those of peer-reviewed journals. medRxiv, a derivative of bioRxiv, provides data sharing through supplementary materials and data/code repositories, with all articles including data sharing statements in the data/code section, though only 27 papers also provide data as supplementary materials. bioRxiv primarily shares

data through supplementary materials, while chemRxiv uses Figshare to support simultaneous uploading and linking of papers and data files. All three platforms automatically assign DOIs and integrate Altmetric social media mentions for early evaluation and accelerated dissemination.

2.3 Data Repositories

For researchers wishing to publish data directly, the most straightforward approach is building dedicated knowledge infrastructure for data storage and sharing. Emerging data repositories can ensure reliable data access and use. Based on the DataCite platform, four repositories were examined: Zenodo, Figshare Academic Research System (Figshare ARS), Mendeley Data, and Harvard Dataverse .

Zenodo, developed under Europe's OpenAIRE program and operated by CERN, stores software code on GitHub. Figshare ARS has partnerships with chemRxiv, Taylor & Francis, Frontiers, and Mendeley, with many journals and preprint platforms supporting storage of supplementary data in Figshare ARS, which can also link to peer-reviewed publications. Harvard Dataverse is Harvard University's institutional repository for research data, with datasets primarily from Harvard's China Data Lab, including COVID case maps, Baidu mobility data, case updates, healthcare facilities, virus reports, Chinese policy and public opinion collection, and GIS and public health research papers and reviews.

All four repositories support DOI assignment, which is critical for data discovery, preservation, and citation. Compared to Harvard Dataverse's complex metadata standards, Figshare ARS and Zenodo adopt minimalist metadata principles for convenient data storage. Figshare ARS and Zenodo provide altmetrics, support multiple file types (text, images, datasets, audiovisual materials, software, interactive resources), and enable upload and download.

2.4 Clinical Trial Registration Platforms

All clinical trials should be registered for ethical and scientific reasons. WHO's International Clinical Trials Registry Platform (ICTRP) certifies eligible national registries and serves as a global one-stop search portal. As shown in Table 5 , ICTRP identified 2,936 COVID-19 clinical trials. The top five registration sources are ClinicalTrials.gov (U.S.), Chinese Clinical Trial Registry (ChiCTR), Iranian Registry of Clinical Trials (IRCT), EU Clinical Trials Register (EUCTR), and German Clinical Trials Register (GermanCTR), demonstrating that Chinese trial registration institutions are actively promoting rapid data sharing alongside international counterparts.

Since 2005, ICMJE has encouraged prospective trial registration, requiring trials to be registered before or at the time of recruiting the first participant as a prerequisite for publication in ICMJE journals. Among 1,443 COVID-19 trials with recruited participants, 686 were prospectively registered, indicating that

while most journals reference ICMJE guidelines, they do not strictly adhere to prospective registration policies.

2.5 Genome/Structure Data Centers

Genetic sequences and protein structure information for pathogens are typically stored in specialized large-scale international data centers, which are crucial for developing diagnostic tests, tracking ongoing outbreaks, and selecting potential interventions. To provide timely COVID-19 information, the China National Center for Bioinformation (CNCB)/National Genomics Data Center (NGDC) established the 2019-nCoV Resource (2019nCoVR), integrating publicly released nucleotide and protein sequence data from major global data centers, including GISAID, NCBI's GenBank, CNGBdb, and the National Microbial Data Center (NMDC) and NCBI/NGDC's Genome Warehouse . Over 30,000 coronavirus sequences involve 541 data submission units, 598 sample collection units, and 456 collection sites.

Increasing numbers of laboratories are releasing COVID-19 genome sequences through GISAID. Researchers can upload extracted viral gene sequences after registration, with each strain receiving a unique identifier and documented collection time, submission date, and submitting laboratory. The earliest sequence was collected on December 23, 2019, and uploaded on January 11 by the Chinese Academy of Medical Sciences to both GISAID and Genome Warehouse. These sequences enable rapid information dissemination through peer-reviewed medical journals for online publication during epidemics. For example, one study obtained eight complete and two partial genome sequences from nine patient samples, deposited in the National Microbiology Data Center (accession NMDC10013002 and genome accession numbers NMDC60013002-01 to NMDC60013002-10), with data from BGI stored in the China National GeneBank (accession CNA0007332-35).

3. Comparison of Data Sharing Mechanisms and Barrier Analysis

If publications are the stars in the scientific universe, data are the "dark matter" of scholarly communication. Recent policies encouraging more open data sharing have diversified implementation pathways. Compared to previous pandemics, COVID-19 data sharing has improved significantly, but remains far from the goal of "open as default, closed as exception." Overall, existing mechanisms fall short of open science's two objectives—transparency and utility—particularly in integrating data from different sources to generate new knowledge and inform decisions more rapidly.

Table 7 analyzes the advantages and disadvantages of five data sharing mechanisms based on three critical attributes: rapid sharing, reusability, and sustainability.

Peer-reviewed journals ensure quality and reliability through peer review, with publisher and journal support facilitating data discovery and access. However, publication is slow, and PDF formats limit machine readability, discoverability, and computational analysis.

Preprints in life and health sciences are increasingly accepted and publish faster than journals, but their reliability is unverified and PDF formats remain non-machine-readable. While preprints accelerate sharing, they face quality assurance challenges, lack indexing in major search engines (Web of Science, Scopus), and researchers worry that preprint sharing may affect publication in high-impact journals, highlighting the tension between sharing speed and quality assurance.

Data repositories use domain-specific metadata standards to ensure discoverability, integrity, and interoperability, offering potentially cost-effective platforms with rich data management functions. However, they have minimal standardized metadata, making data difficult to discover, and their long-term preservation and sustainability remain unproven.

Clinical trial registration platforms and **genome/structure data centers** have the highest data standardization and reusability due to strong professional community support. These international or national data sharing infrastructures require substantial financial and human resources. For example, in FY2016, the U.S. National Library of Medicine allocated $190 million to process and provide massive data from novel sequencing, microarray, and small-molecule screening technologies, plus clinical trial data. NCBI employs 288 full-time staff to manage and maintain GenBank. High operational costs mean most genome sequencing occurs in high-income countries, potentially disadvantaging researchers in low-income nations.

Both journals and preprints can enhance research transparency by sharing data, but face difficulties addressing the higher-level goal of data reusability. As the global data ecosystem shifts toward discipline-agnostic, general-purpose repositories, most adopt minimal metadata standards (e.g., Dublin Core) for convenient storage and interoperability. In this study, Figshare ARS shared the most COVID-19 datasets, using DataCite metadata standards for automatic integrity checks, Creative Commons compliance, and collaboration with academic publishers and institutions. Funding for data repositories is complex: Figshare ARS is supported by Holtzbrinck Publishing Group, while Mendeley was acquired by Elsevier and backs up all data to DANS, a Dutch government-supported archiving service. Harvard Dataverse relies on Harvard University funding. The long-term sustainability of repository platforms requires further validation.

In summary, major barriers to rapid data sharing during public health emergencies include: (1) inadequate data sharing infrastructure such as data centers and repositories; (2) lack of clear incentives for data attribution and academic recognition for providers and users, with limited emergency-specific motivation for re-

searchers; (3) potential economic losses from commercialization and intellectual property; (4) inequalities between high- and low-income countries in capacity and funding, with varying costs and levels of data management system adoption; (5) academic reputation risks from premature data sharing; and (6) regulatory and governance issues related to privacy and informed consent in experimental treatments and clinical care, plus compliance with national and international ethical and legal requirements. These barriers correspond to six categories—technical, motivational, economic, political, legal, and ethical factors—with the first three being most critical. Their complex interrelationships and feedback mechanisms mean single-factor solutions are unlikely to succeed.

4. Recommendations for Enhancing China's Public Health Emergency Data Sharing Capacity

During the early COVID-19 outbreak, China was the first to release viral genome sequences and related information to the world, demonstrating transparency and responsibility. However, incidents also occurred where researchers delayed disclosure due to publication priority concerns or had their data used by others to publish first. In reality, global public health data sharing frameworks have emerged from crises rather than proactive planning. Given vast differences in social, political, cultural, and legal systems, mandatory, one-size-fits-all global mechanisms often prove ineffective. Based on historical experience, China should enhance its public health emergency data sharing capacity by following global principles while addressing local needs and governance characteristics through strategic planning, infrastructure, stakeholder coordination, and ethical-legal frameworks.

4.1 Establish Strategic Priority for Data Sharing in National Public Health Emergency Preparedness and Response

In April 2018, China's State Council issued the Scientific Data Management Measures, establishing national policy principles for improving data openness and sharing for the first time. In WHO's COVID-19 research roadmap, data sharing is listed as a core priority task involving needs assessment, target setting, coordinated implementation of targeted interventions, and monitoring and evaluation. Specifically, China should develop an implementation roadmap for data sharing that clarifies data collection purposes, user needs, and the ecosystem of actors involved in data collection, management, analysis, and use. This should inform information management systems and procedures needed before, during, and after public health events to raise awareness and influence behavior among decision-makers and the public.

4.2 Develop and Build Global Data Sharing Infrastructure Across Regions, Disciplines, and Media

International consensus on sharing raw data in principle is insufficient. Research data are the "oil" of future scientific innovation and a critical foundation for national population health and biosecurity. Current data sharing platforms and authoritative journals are predominantly owned by technologically leading countries (U.S., UK) and international research institutions. To remain competitive, China should first develop data sharing platforms for diseases with epidemic potential, including creating open-access journals, preprint services, clinical trial registries, and genome/structure databases to support international and domestic research collaboration. Second, China should build general-purpose data repository platforms, clarify data standardization methods, develop standardized tools, and simplify data governance processes to ensure data quality and address technical challenges in implementing data sharing protocols. Finally, China should integrate social media, traditional media, and other rapid sharing channels to disseminate and evaluate verified accurate information.

4.3 Organize and Coordinate Data Sharing Stakeholders in Public Health Emergency Risks

Key stakeholders can be divided into three groups: data providers, data users, and data sharing facilitators. Responsible stakeholders must have the willingness, skills, and capacity to share data and integrate them into decision-making. Overcoming barriers requires focusing on coordination between institutions and actors in two main areas: First, researchers are often both providers and users, requiring reasonable incentives for data attribution and academic recognition. Training and outreach materials should help researchers understand the benefits of data sharing during emergencies to enhance transparency and promote a data sharing culture. Second, facilitators include policymakers from government and funding agencies, journal and database platform providers, and libraries and universities. Government, funding, and university administrators should promote funding policies that recognize research quality and public health impact. Journals and database providers must shift business models to incentivize timely data sharing rather than prioritizing publication. Libraries should integrate into the entire research process, providing sharing services tailored to different data needs and developing new service models.

4.4 Strengthen Research on Ethical and Legal Issues in Data Sharing

Transparent data sharing during emergencies can enhance scientific value, and data producers and publishers have ethical responsibilities to collect and share data through appropriate informed consent procedures and privacy protection to facilitate rapid public health data dissemination. Researchers face pressure to conduct original research and publish in high-impact journals, while corporate sponsors pursue commercial interests. Therefore, ethical responsibilities of researchers, sponsors, public health personnel, and publishers regarding data

management and sharing must be carefully considered in relation to the broader research community, health and government agencies, and the public. Given the potential severity and time sensitivity of emergencies, legal responsibilities for shared data must be clarified to enable timely public health measures. According to WHO's COVID-19 research roadmap, data ethics and data sharing are closely related interdisciplinary research priorities. This paper does not elaborate further on data ethics and law, which will be addressed in future work.

References

[1] Pisani E, Ghatature A, Merson L. Data sharing in public health emergencies [EB/OL]. [2020-03-10]. https://www.glopid-r.org/wp-content/uploads/2017/02/data-sharing-in-public-health-emergencies-case-studies-workshop-reportv2.pdf.

[2] Data sharing in public health emergencies: learning lessons from past outbreaks [EB/OL]. [2020-03-10]. https://www.glopid-r.org/wp-content/uploads/2017/02/data-sharing-in-public-health-emergencies-case-studies-workshop-reportv2.pdf.

[3] Enserink M. Calling all coronavirus researchers: keep sharing, stay open [J]. *Nature*, 2020, 578(7793): 7.

[4] Yozwiak N, Schaffner SF, Sabeti PC. Data sharing: make outbreak research open access [J]. *Nature*, 2015, 518(7540): 477-479.

[5] Boosting access to disease data [J]. *Nature*, 2006, 442(7106): 957-957.

[6] Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data—from vision to reality [J]. *Eurosurveillance*, 2017, 22(13): 30494.

[7] Littler K, Boon WM, Carson G, et al. Progress in promoting data sharing in public health emergencies [J]. *Bulletin of the World Health Organization*, 2017, 95(4): 243.

[8] Wellcome Trust. Statement on data sharing in public health emergencies [EB/OL]. [2020-03-10]. https://wellcome.ac.uk/press-release/statement-data-sharing-public-health-emergencies.

[9] Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors (ICMJE) [J]. *Annals of Internal Medicine*, 2016, 164(7): 505-506.

[10] A guide to sharing the data and benefits of public health surveillance [EB/OL]. [2020-03-20]. https://www.chathamhouse.org/sites/default/files/publications/research/2017-05-25-data-sharing-guide.pdf.

[11] Moorthy V, Restrepo AMH, Preziosi MP, et al. Data sharing during the novel coronavirus public health emergency of international concern [J]. *Bulletin of the World Health Organization*, 2020, 98(3): 1-3.

[12] Sharing research data and findings relevant to the novel coronavirus (nCoV) outbreak [EB/OL]. [2020-03-01]. https://wellcome.ac.uk/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-covid-19-outbreak.

[13] COVID-19 Public Health Emergency of International Concern (PHEIC) global research and innovation forum: towards a research roadmap [EB/OL]. [2020-03-04]. https://www.who.int/blueprint/priority-diseases/key-action/Global_{{{Research}}}*{{Forum}}*{{{FINAL}}}*{{VERSION}}*{{{for}}}*{{web}}*{{{14}}}*{{feb}}*{{{202

[14] Pradhan P, Pandey AK, Mishra A, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag [EB/OL]. [2020-03-01]. https://www.biorxiv.org/content/10.1101/2020.01.30.927871v1.full.pdf.

[15] Hardenman M. Coronavirus research on figshare [EB/OL]. [2020-03-13]. https://doi.org/10.6084/m9.figshare.c.4836594.v8.

[16] Gopala D, Wallace JD, Aminawung JA, et al. Adherence to the International Committee of Medical Journal Editors' (ICMJE) prospective registration policy and implications for outcome integrity: a cross-sectional analysis of trials published in high-impact specialty society journals [J]. *Trials*, 2018, 19(1): 1-13.

[17] Zhao Wenming, Song Shuhui, Chen Meili, et al. 2019 novel coronavirus information database [J]. *Hereditas*, 2020, 42(2): 212-221.

[18] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding [J]. *The Lancet*, 2020, 395(10224): 565-574.

[19] Borgman CL. *Big Data, Little Data, No Data: Scholarship in the Networked World* [M]. Translated by Meng Xiaofeng, Zhang Xiang, Zhao Erping, et al. Beijing: China Machine Press, 2017.

Author Contributions: Cui Yuhong: Research framework design, data analysis, and paper writing;
Wang Sa: Data collection and policy recommendations section writing.

Note: Figure translations are in progress. See original paper for figures.