**Managing Future Pandemics: Benefits and Challenges of Creating a Common Data Space for Highly-Infectious Diseases**

By Stephanie H. Chin and Caitlin T. Chin

*Background*

The COVID-19 pandemic calls to attention that—despite best intentions and efforts—real-time data emerging from global crises may be uncertain, rapidly evolving, incomplete, or even misleading. For example, the lag between COVID-19 transmission and the onset of symptoms, as well as the lag between getting tested and receiving test results, can lead to outdated infection rate estimates and dynamically-changing public health guidance—which in turn reduces public understanding and compliance. Governments and academic researchers must choose whether and how to update backlogged information or retroactively fix past statistics, which may lead to changing, reversing, or delayed policies.

During recent epidemics, including the COVID-19 pandemic, sharing genomic sequencing data in public databases and data repositories such as GenBank and GISAID[1] has proven extremely valuable—and is supported by international agreements such as the 1996 Bermuda Principles and 2010 Nagoya Protocol. Other calls for public health data sharing[2] have led to the development of innovative platforms to map disease occurrence,[3] leverage open source and social media intelligence about public health,[4] and even collect data through participatory crowdsourcing.[5] However, these efforts are fragmented.

*Common Data Spaces*

The establishment of a common data space for highly-infectious diseases would offer the potential to improve data sharing during global public health crises. Common data spaces integrate multiple data sources, enabling a more comprehensive analysis of data based on greater volume, range, and access. At its essence, a common data space is like a public library system, which has collections of different types of resources from books to video games; processes to integrate new resources and to borrow resources

---

[1] Global Initiative on Sharing All Influenza Data (GISAID), https://www.gisaid.org/.

[2] Jane Qiu, "One world, one health: combating infectious diseases in the age of globalization," *National Science Review,* Volume 4, Issue 3, May 2, 2017, pp. 493-499. https://academic.oup.com/nsr/article/4/3/493/3789515;
Rose Bernard, et al., "Intelligence and global health: assessing the role of open source and social media intelligence analysis in infectious disease outbreaks," *Journal of Public Health* 26, 509-514, 2018. https://link.springer.com/article/10.1007/s10389-018-0899-3;
Nathan Yozwiak, et al., "Data sharing: Make outbreak research open access," *Nature,* February 25, 2015, https://www.nature.com/news/data-sharing-make-outbreak-research-open-access-1.16966;
David Bloom and Daniel Cadarette, "Infectious disease threats in the twenty-first century: Strengthening the global response," Frontiers in Immunology., March 28, 2019, https://www.frontiersin.org/articles/10.3389/fimmu.2019.00549/full;
Simon Hay, et al., "Big data opportunities for global infectious disease surveillance," PLOS Medicine, April 2, 2013, https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001413#s4.

[3] *HealthMap.* Retrieved from http://www.healthmap.org.

[4] Rose Bernard, et al., "Intelligence and global health: assessing the role of open source and social media intelligence analysis in infectious disease outbreaks," *Journal of Public Health* 26, 509-514, 2018. https://link.springer.com/article/10.1007/s10389-018-0899-3.

[5] Simon Hay, et al., "Big data opportunities for global infectious disease surveillance," PLOS Medicine, April 2, 2013, https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001413#s4;
*FluNearYou.* Retrieved from https://flunearyou.org/#!/;
*Influenzanet.* Retrieved from http://www.influenzanet.eu/.

from other libraries; a catalog system to organize, sort, and search through resources; a library card system to manage users and authorization; and even curated collections or displays that highlight themes among resources.

Even before the pandemic, there was significant momentum to make critical data more widely accessible. In the United States, Title II of the Foundations for Evidence-Based Policymaking Act of 2018, or the OPEN Government Data Act,[6] requires federal agencies to publish their information online as open data, using standardized, machine-readable data formats. This information is now available on the federal data.gov catalog and includes 50 state- or regional-level data hubs and 47 city- or county-level data hubs.[7] In Europe, the European Commission released a data strategy in February 2020 that calls for common data spaces in nine sectors, including healthcare, shared by EU businesses and governments.[8]

Going further, a global common data space could help identify outbreaks and accelerate the development of new treatments by compiling line list incidence data, epidemiological information and models, genome and protein sequencing, testing protocols, results of clinical trials, passive environmental monitoring data, and more.

Moreover, it could foster a common understanding and consensus around the facts, which is a prerequisite to reach international buy-in on policies to address situations unique to COVID-19 and other pandemics, such as the distribution of medical equipment and PPE, disruption of the international tourism industry and global supply chains, social distancing or quarantine, and mass closures of businesses.[9]

*Challenges*

Yet, setting up a common data space that is usable and secure is no simple task. Despite widespread consensus within academia on the importance of sharing public health data, there are real-world technical, geopolitical, and ethical barriers to implementation.

First is the technical challenge of setting up a comprehensive, secure, and usable data space system. Integrating data from multiple data sources can be time-consuming and difficult, especially considering low data quality, disparate methods of data collection, lags in data reporting, and inherent uncertainties. Thus, it is important to regularly audit data in shared data spaces—flagging poor data quality and outdated information—to help communicate levels of confidence or uncertainty in the data. In addition, emerging AI and ML algorithms and data standards could help automatically provide basic functionality, enabling researchers to focus their efforts on more advanced integrations and operations.[10] The application of blockchain could help ensure systems are secure and resilient against accidental and malicious data corruption.

---

[6] OPEN Government Data Act of 2018, Pub. L. 115-435. https://www.congress.gov/bill/115th-congress/house-bill/4174/text#toc-H8E449FBAEFA34E45A6F1F20EFB13ED95.

[7] Open Government. Retrieved from https://www.data.gov/open-gov/ on September 10, 2020.

[8] *A European strategy for data.* Retrieved from https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf.

[9] *Stasis theory.* Purdue University Online Writing Lab. Retrieved from https://owl.purdue.edu/owl/general_writing/the_writing_process/stasis_theory/index.html.

[10] This "data space" approach helps avoid the high up-front costs of cleaning, processing, and setting up the data ex ante.

On the geopolitical front, issues of data protectionism, national security, economic competition, lack of trust, and differing privacy regulations and values impede the development of a common data space. Pre-print publication policies have helped to incentivize data sharing and safeguard academics' concerns about IP, data ownership, and publication rights, yet there remains a gap in translating academic research to policymakers and the general public.[11] In the past, the exchange of epidemiological prediction models, risk maps, and disaster planning simulations has helped researchers understand country-specific concerns while navigating future uncertainty and high levels of stakeholder complexity.[12]

Underlying these geopolitical issues are ethical questions about data access, equity, and privacy. For instance, how can we ensure that the costs and benefits of a common data space are fairly distributed? It will become necessary to mitigate gaps in disease detection in under-resourced areas, while simultaneously ensuring fair access among communities who contribute data to resulting medicines and treatments.[13] On the other hand, it is also essential to consider how to address "free riding" nations who benefit from the common data space without sharing their own data. Other issues include how to address ownership and attribution when researchers share data, what ethical research and accountability standards are necessary, under which contexts to require informed consent from research participants, and how to offer widely-accessible data in public health emergencies while maintaining the privacy of all individuals involved.[14]

With these considerations in mind, the question then turns to whether a national or regional system of data sharing is the most realistic goal—or whether it is possible to achieve a truly global system of common data sharing. The irony of the situation may be that common data spaces may help increase international trust and cooperation during future pandemics—but international trust and cooperation are also required to establish them in the first place.

[11] Caroline Buckee, et al., "Productive disruption: opportunities and challenges for innovation in infectious disease surveillance," BMJ Global Health, 2018, https://gh.bmj.com/content/3/1/e000538.
[12] Samuel Brannen and Kathleen Hicks, "We predicted a coronavirus pandemic. Here's what policymakers could have seen coming," *Politico,* March 7, 2020, https://www.politico.com/news/magazine/2020/03/07/coronavirus-epidemic-prediction-policy-advice-121172.
[13] "Global network of scientists work to track COVID-19's spread," *The World,* May 1, 2020, https://www.pri.org/file/2020-06-26/global-network-scientists-work-track-covid-19-s-spread.
[14] Sam Halabi, et al., "The law and ethics of data sharing during infectious disease emergencies," *Journal of Health Care Law & Policy, Forthcoming,* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375089.