ATTRIBUTIVE JUSTICE IN INTERNATIONAL LAW: THE GLOBAL LAW AND INFRASTRUCTURE OF PATHOGEN GENOMIC SEQUENCE DATA-SHARING AND BENEFIT-SHARING

BENEDICT KINGSBURY*

A succession of epidemic diseases among humans in the first decades of the 21st century renewed long-standing controversies about power imbalances and justice in the global production, use, and distribution of scientific data and its benefits. A new area of contention concerns digital genomic sequence data (GSD). The largely-forgotten idea of 'attributive justice', articulated by Hugo Grotius (1625), helps make sense of otherwise-disparate demands for GSD justice.

At least two kinds of attributive justice claims are made in relation to GSD. One is for attribution of credit to scientists and others involved in medical services or other procurement of samples—a scientist-focused attributive justice. These claims are mobilized especially in efforts to rectify existing power and resource imbalances in science production, both within national societies and by scientists from developing country. These claims have considerable traction, but not in formal international law.

A second claim relates to demands by developing countries either to control GSD, or at least to receive benefits from commercial use of it when the underlying biological sample originates specifically in their territory. These claims have been pursued in efforts to extend the 2010 Nagoya Protocol. Other existing or pending international treaty regimes embedded in entirely separate institutions also address benefit sharing in relation to oceanic, plant, or human digital sequence sharing, complicating the formation of a coherent or unified set of rules. Contentions about widely used sets of data-governance principles such as Findable, Accessible, Interoperable, and Reusable (FAIR) data also arise in each treaty regime.

Infrastructural regimes for the sharing of sequences have become major sites for both the scientist-relative and state-relative attributive justice claims. The most widely used platform for access to GSD of all kinds is the International Nucleotide Sequence Database Collaboration (INSDC) (including GenBank). A leading alternative is GISAID, which is similar in being free to use but conditions GSD access and sets requirements for attribution of

* Director, Institute for International Law and Justice, NYU Law School. The author collaborates closely with Angelina Fisher and Thomas Streinz in work on global data law and infrastructures, and is deeply grateful to them, along with Andrew Hurrell, Nahuel Maisley, Alejandro Rodiles, Yirong Sun, and the late Sally Engle Merry. Thanks also to participants in workshops with the Berlin KFG research group and at McGill University Faculty of Law, to Julia Spencer and Daniel Sive in the IILJ, and to the excellent *NYU Journal of International Law and Politics* editors.

scientific credit. While GISAID goes further than INSDC in supporting scientist-relative attributive justice claims, the two infrastructures are broadly similar with regard to state-relative claims for attribution of GSD and benefit-sharing. The infrastructures have recently begun trying to ensure that metadata accompanying each sequence attributes it to samples taken from a particular country, but not that the GSD is systematically linked to its commercial outcomes. These infrastructures embed norms and ideologies of their original builders, such as a normative commitment to 'Open Science', and the economic and epidemic-security interests of richer OECD countries.

Attributive justice entitlements of particular scientists and states will not leverage universal principles of distributive virus- and vaccine-justice but are reinforcing significant shifts toward orders of respect and recognition in global health research and (slowly) in sequencing infrastructures. The contributions of attributive justice have been underestimated.

I. Introduction

Digital genomic sequence data (GSD) from pathogen samples plays a vital role in diagnostic, therapeutic, immunizational, and epidemiological responses to both novel and known pathogens with severe epidemic effects or potential. A single complete sequence may have great value for some purposes—the first COVID-19 diagnostic tests were developed in Germany in early 2020 based simply on the original Wuhan SARS-CoV-2 virus sequences prepared and uploaded in China.[1] The nucleotide sequence data is often more useful when accompanied by metadata about the context of the sam-

1. Mathieu Gand et al., *Use of Whole Genome Sequencing Data for a First in Silico Specificity Evaluation of the RT-qPCR Assays Used for SARS-CoV-2 Detection*, 21 (15) Int'l J. Molecular Sci. (2020), at 3; Lyle Fearnley, *Viral Sovereignty*

ple including information about the patient, exposure, symptoms, and medical trajectories as well as sample-handling information.[2] For many purposes, analysis of one digital sequence may depend on the availability of sequences from other samples and metadata accompanying those sequences.[3] When all of this digital data is available, a great deal of scientific research, academic publication, and product development proceeds with no need for direct access to any physical sample. The work is conducted "*in silico*"—using computing-intensive analytic services provided by the infrastructures or cloud or super-computing companies,[4] and using and contributing to synthetic biology, synthetic data, artificial intelligence (AI) models, and other fields.[5]

The availability of GSD (including metadata) depends on an intricate series of links from health surveillance and sample collection to sequence production, and then on the willingness of those who have the GSD to share it, and the choices they make among different GSD-sharing infrastructures, each of which imposes conditions and supplies some degree of curation. This article examines the interactions of justice claims with some of the regulatory and infrastructural issues that arise along this path.

---

or Sequence Etiquette?, 14 EAST ASIAN SCO., TECH. AND SOC.: AN INT'L J. 479, 482 (2020).

2. Vivien G. Dugan et al., *Standardized Metadata for Human Pathogen/Vector Genomic Sequences*, 9 PLOS ONE e99979 (2014), at 4-7 (reporting the development of pertinent metadata standards).

3. *See Synthesis Of Views And Information Related To Digital Sequence Information On Genetic Resources*, CBD/DSI/AHTEG/2020/1/2, at 10–11 (Mar. 3, 2020) (describing norms, practices and benefits of open access in digital sequence information); and *see generally* Animesh Ray, *Machine learning in postgenomic biology and personalized medicine*, 12 WIREs DATA MINING AND KNOWLEDGE DISCOVERY 1451 (2022) (describing uses of machine learning applications based on large datasets in biology).

4. Masanori Arita et al., *The International Nucleotide Sequence Database Collaboration,* 49 NUCLEIC ACIDS RSCH. D121, D122–D123 (2021).

5. *See generally* WORLD HEALTH ORG., GLOBAL VACCINE MARKET REPORT (2022) (presenting context and an analysis of vaccine markets including available price and procurement data); *see also* Michelle Rourke, *Access and Benefit-Sharing DNA Componentry for Plant Synthetic Biology: Bioparts Expressed in Plant Chassis*, 4 PLANTS, PEOPLE, PLANET 75 (2022) (on how benefit-sharing rules may affect deployment of synthetic biology technologies which fragment genetic resources into small interchangeable bio-parts).

An open letter in 2021, signed by nearly 800 molecular-genomic scientists mainly in prosperous countries, called upon researchers and policymakers to submit all SARS-CoV-2 genomic sequence data to the fully open and free-of-charge International Nucleotide Sequence Database Collaboration (INSDC). The letter stated: "As much research and healthcare data as possible need to be taken out of silos and stored into an open, connected and FAIR (Findable, Accessible, Interoperable and Reusable) environment" and "not unnecessarily siloed, fragmented and closed."[6] Many of the arguments it made for submitting sequence data to the INSDC databases were infrastructural: the databases "ensure the rapid dissemination of data with maximal impact due to their connectivity to the global bioinformatics data infrastructure."[7] The scientists urging use of INSDC infrastructure also emphasized that the "INSDC quickly makes submitted data freely and permanently available to everyone, without the need to log in or any restrictions on reuse."[8] On this point, the letter represented a veiled challenge to GISAID, another major free-of-charge public sequence repository, which from the beginning of the COVID-19 epidemic had become by a significant margin the largest repository of SARS-CoV-2 sequences.[9] Anyone may join GISAID, but a real ID and log-in are required to access or use data,[10] and users pledge that any publication they make using the data will acknowledge also the originating (sample-collecting) and submitting (sequencing) laboratories which produced the data, and that they will make best efforts to collaborate with these in research or publications relating to the data.[11] The reasons why GISAID was favored by some scien-

---

6. Edith Heard et al., *Open Letter: Support Data Sharing for COVID-19* (Jan. 29, 2021), https://www.covid19dataportal.org/support-data-sharing-covid19 [https://perma.cc/G5LJ-YT9B]; Richard van Noorden, *Scientists Call for Fully Open Sharing of Coronavirus Genome Data*, 590 NATURE 195 (2021).

7. Edith Heard et al., *supra* note 6.

8. *Id.*

9. Amy Maxmen, *One Million Coronavirus Sequences: Popular Genome Site Hits Mega Milestone*, NATURE NEWS (Apr. 23, 2021), https://www.nature.com/articles/d41586-021-01069-w [https://perma.cc/PB3R-XBE5].

10. GISAID, https://gisaid.org/help/faq/ [https://perma.cc/7RRM-REFY] (last visited Apr. 24, 2023).

11. *Citation and Acknowledgement Guide*, GISAID, https://gisaid.org/publish/ [https://perma.cc/YX5S-5H9J] (last visited Apr. 24, 2023). The same

tists, including many based in developing countries, was indicated by a South African editorial responding to the open letter: "a push from wealthy countries for open data is suspect, given how often scientists in the global south go unacknowledged."[12] Rolf Apweiler, co-director of the European Bioinformatics Institute (EBI) in the United Kingdom and one of the originators of the open letter, responded that: "The focus on low- and middle-income countries is bizarre because their amount of data is relatively little."[13]

In parallel to this debate about scientist-relative justice—in particular, inequitable lack of recognition and resourcing of scientists and other contributors in developing countries—was a wider state-relative struggle about flows of resources to developing countries in recognition of their contributions of genetic resources and sequences used in the effort to develop commercial-scale tests, therapeutics, and vaccines to combat disease or achieve other benefits.

The claim for scientist-attribution is what this article terms an attributive justice claim—but it will also have distributive results in that resources and status are likely to flow when those doing the work are recognized for it. The state claims are also formally for attributive justice— attributing genetic resources to particular states so that commercial benefits from the resources flow back to them—but their underlying motivation is also distributive, in seeking, in aggregate, to increase benefit flows to countries with limited prosperity and great needs. The scientist-relative attributive justice claims are increasingly supported as an efficient and equitable means to as-

---

webpage has the banner phrase: "Acknowledging contributors is not a choice, but an essential requirement."

12. Opinion, *Data Imperialism in a Pandemic: How to Make Things Fall Apart*, IOL, https://www.iol.co.za/news/opinion/data-imperialism-in-a-pandemic-how-to-make-things-fall-apart-f707fe96-fc6f-4866-aa9f-7c0aaa1f97db [https://perma.cc/TEZ2-6JVJ] (May 4, 2021), *quoted* in Amy Maxmen, *The Flip Side of Unrestricted Viral Genome Sequencing*, 593 NATURE 176, 177 (May 13, 2021).

13. Maxmen, *supra* note 12, at 177. Whether this quotation is accurate or taken out of context is not known to the present author. In her *Nature* report on these equity issues, Amy Maxmen noted that by May 2021 all of Africa had uploaded around 13,000 sequences to GISAID, and South America had uploaded 14,000 sequences, for instance, compared with about 380,000 from the United Kingdom alone. She noted too that only 0.3% of then-supplied COVID-19 vaccination doses had gone to developing countries.

sure equality of treatment and a better distribution of power and resources in global science, and, hence, also to improve global science as a whole. The state-relative attributive justice claims, however, have been cumbersome and inefficient to administer, not congruent with massive multi-source uses of GSD in biotech, and not reliably aligned with distributive justice insofar as benefits would flow based on the location from which an original sample came rather than need or future capacity-building.

The two struggles are different in these respects as well as in the legal and other norms and institutions that are invoked. The struggles are intertwined, however, in that sequencing and the circulation of sequence data are the touchpoint for both scientist-recognition and state-benefit claims. The effective pursuit of each category of claims depends in part on the data-sharing infrastructures through which the sequence data is made available, making these infrastructures (perhaps unexpectedly) major focal points in these debates.[14] These infrastructures (GISAID, and the INSDC and its three member entities) are difficult for most states or scientists from developing countries to regulate,[15] but key funders and journals and groups of scientists may exert persuasive or even decisive influence on policies of the infrastructures—and, in extremis, new entrants could enter to compete with the established infrastructures. Conversely, the infrastructures, when buttressed by other entities such as leading science funders and academic journals, may themselves exercise regulative power in imposing their conditions on scientists wishing to upload to or make use of the sequence databases. This case thus invites exploration of the complex interactions between regulation of infrastructure and infrastructure-as-regulation—a set of interac-

14. *See* RESEARCH DATA ALLIANCE COVID-19 WORKING GROUP, RECOMMENDATIONS AND GUIDELINES ON DATA SHARING (June 30, 2020) https://zenodo.org/record/3932953#.Xx_f7Z4za70 [https://perma.cc/7L9M-7UQH] (providing guidance on data sharing and highlights the importance of data sharing in bringing rapid results during public health emergencies); *see also* WORLD HEALTH ORGANIZATION, WHO GUIDING PRINCIPLES FOR PATHOGEN GENOME DATA SHARING (2022), https://www.who.int/publications/i/item/9789240061743 [https://perma.cc/WSY2-FJP4] (encouraging data sharing of pathogen genome data to protect global public health).

15. *See* Gian Luca Burci and Frederic Perron-Welch, *International Sharing of Human Pathogens to Promote Global Health Security—Still a Work in Progress*, 25:13 ASIL INSIGHTS 1 (2021) (discussing GISAID and its free user access).

tions encapsulated in the premises of the academic "Infrareg" project.[16]

This struggle over infrastructures and rules—while internecine and subordinate to strong beliefs among scientists worldwide in the value of open flows of scientific information to combat pathogens—brought again to the fore some fundamental and longstanding issues about power, knowledge, and infrastructure in the global sharing of scientific data. The divergences between the INSDC and GISAID on scientist-attribution requirements are, in part, differences in norms of attributive justice and the significance accorded to such justice claims within each of these infrastructural configurations. These are linked to power relations, status, hierarchies, and recognition.

The term "attributive justice" was used by Hugo Grotius in *The Rights of War and Peace* (1625); he was updating Aristotle's terminologies of justice and developing rules and principles which he thought could apply anywhere and across different cultures and circumstances. "Attributive justice" is not a standard term, and even Grotius did not use it very often. This article re-introduces it as part of an effort to better capture one idea of justice that is present within many kinds of organized relations but not yet fully elucidated in standard contemporary discussions of justice. The idea of attributive justice seems to do some animating work in each of the two kinds of GSD claims discussed here. Certainly, the scientist-relative claims draw mainly on ethical norms of science, and the state-relative claims draw on an idea that the sovereignty of the state entails a right to control access and to condition use of national resources on the provision of benefits to its people. However, an appeal to justice is imbricated in these claims beyond what is captured in these sources of right the claims draw on. For a scientist, or a lab worker or health worker, there is an appeal to a particular kind of justice in saying, "I did that work; my creativity and effort ought to be acknowledged and attributed to me, and least where it is infrastructurally possible and reasonably practicable." This is not reducible simply to intel-

16. Benedict Kingsbury, *Infrastructure and InfraReg: On Rousing the International "Wizard of Is,"* 8 CAMBRIDGE INT'L L.J. 171 (2019); Benedict Kingsbury, *Introduction to the Symposium on Infrastructuring International Law*, 117 AJIL UNBOUND 1 (2023). This is the animating idea of the "InfraReg" research project at NYU Law School (iilj.org/infrareg) [https://perma.cc/4S3S-84X3].

lectual property rights to share in returns, or labor rights to be paid for work done. The state-relative claims are indirect claims the government or another entity makes for justice on behalf of people whose traditional knowledge or care and nurturing of biodiversity is to be acknowledged, and they may also be claims on behalf of the more-than-human life, ecosystem, and place that have enabled these particular life-forms, and which themselves must continue to flourish and adapt.[17] What is sought in each case is an attribution, a recognition of creation and work and contribution in some way embodied in the sequenced sample and the sequence itself. Hence the term 'attributive justice'. In its outward presentation, it is a claim to attribution as a kind of recognition that might bring with it some claims to participate in ensuing projects or in governance of them, and to object to misrecognition.[18] These justice claims are imperfect; they do not in themselves determine that particular others are under specific duties. Efforts are made to actualize them by establishing such duties elsewhere, whether in positive law, guiding principles, ethical practices, or infrastructural rules. Such efforts are met by countervailing claims of rights or justice, or competing policy prescriptions, and in any case confront divergent interests as well as the limits of

17. Claims made in relation to the neem tree and products derived from it have been examples since the 1980s, when a wide range of groups and eventually the Indian government became mobilized against 'biopiracy'. David Dickson and K. S. Jayaraman, *Aid Groups Back Challenge to Neem Patents*, 377 NATURE 95 (1995) (on challenges by development and environmental groups to patents issued in Europe and the United States). *See also* Ravinda N. Kharwar et al, *Harnessing the Phytotherapeutic Treasure Troves of the Ancient Medicinal Plant Azadirachta indica (Neem) and Associated Endophytic Microorganisms*, 86 PLANTA MED 906 (2020) (noting compounds isolated from neem with known or potential antimicrobial, antimalarial and other medicinal properties, as well as compounds from endophytes found in neem trees). The multi-valence of *Neem* claims and the historical and power assertions linked to them involve complex social and justice issues within Indian society and politics, and throughout South Asia, that are not discussed here.

18. This has some affinity with explorations in Jeremy Waldron, 'Contributive Justice' (Feb. 2023, draft on file with author). He builds from and beyond Michael Sandel's argument that workers may seek "a greater measure of *contributive* justice—an opportunity to win the social recognition and esteem that goes with producing what others need and value." Michael Sandel, *What Liberals Get Wrong About Work*, THE ATLANTIC, Sept. 2, 2020, https://www.theatlantic.com/ideas/archive/2020/09/contributive-justice-and-dignity-work/615919/ [https://perma.cc/Y5K2-M4ZK].

what is practicable and affordable. Thus, the value of attributive justice might be high in one case (or in one category of claims) but much less compelling in another. These efforts and their challenges will be discussed in this article. The efforts and institutions involved span the world, so inevitably are bound up with historic injustices, vastly discrepant resources, embedded structures of power and status inequality, and demands for rectification and change. The attributive claims thus have running underneath them a powerful current of claims for compensatory or distributive justice, always present but often not on the surface of the discourse until a particular event, such as an epidemic, cracks the carapace. Grotius regarded attributive justice as pertaining to and fostering virtues in society that are beneficial to others, such as liberality, mercy, or prudent administration.[19] But it was an imperfect Right, not imposing counterpart obligations on others (in contrast to the perfect Right to get back one's goods from another who wrongfully possesses them).[20] To many modern readers, distributive justice (the Aristotelian term) seems to disappear from significance in Grotius's tome when he replaces it with attributive justice, and indeed, both distributive justice and also what Grotius calls "attributive justice" end up with only moral suasion and, thus, a very limited place in politics.[21] Others, however, see Grotius's wider idea of justice as vital in overcoming the limits of "strict" justice and offsetting their imbalances,[22] and also in his means of bringing into natural law and the law of nations a responsibility for the good of others and for the common good.[23] Institutions that are just may be means for building orders of respect and recognition, and

19. HUGO GROTIUS, THE RIGHTS OF WAR AND PEACE 142 (2005) (Bk I, ch. i, s. 7) [1625].

20. *Id.* (stating that under natural law a poor person, however deserving of Alms, has no right to them being given, but if given Alms, the recipient has an absolute right to keep them and demand their restoration if then the donor purports to take them back.)

21. *See* RICHARD TUCK, THE RIGHTS OF WAR AND PEACE 98–99 (1999) (describing the differences between corrective and distributive justice). In this view, it is only strict justice (including through war) that will serve to overcome the dangerous use of force.

22. Jeremy Seth Geddert, *Beyond Strict Justice: Hugo Grotius on Punishment and Natural Right(s)*,76 REV. POL. 559, 563–4 (2014).

23. Janne Nijman, *Grotius' 'Rule of Law' and the Human Sense of Justice: An Afterword to Martti Koskenniemi's Foreword*, 30 EJIL 1105, 1111 (2020).

thereby building trust and reciprocity as well as dissipating resentment.[24] The attributive claims relating to virus genomic sequence data might be claims of this sort. In practical politics they have remained quite distant from the imperative claims to vaccine distribution and health-care access, claims which tend instead to be pursued through human rights and health equity arguments, as well as arguments based on shared self-interest.[25] Increasingly, however, the attributive justice claims with regard to pathogen data have been a site for, and have converged with, major changes in science ethics and practice with regard to developing countries and indigenous or minority communities,[26] including important commitments to community control and benefit flows.[27] The sequence data infrastructures and surrounding science funding and publishing ecosystems have adopted regulatory requirements consonant with these shifts, and begun to redefine what is meant normatively by publicness in these infrastructures. These claims will be developed in the remaining sections of this article.

Genomic sequencing and the sharing of sequence data on a significant scale dates to the 1960s, with sharing initially done mainly by print-format publications, then electronically as computer power and networking increased in the 1970s, and consolidated in the establishment in 1992 of the Europe-Japan-US International Nucleotide Sequence Database Collaboration (INSDC)[28] including the NIH-supported GenBank.[29]

---

24. *See id.* at 1113 (2020) (referring also to the thought of Paul Ricoeur and his theory of "small ethics").

25. *See* United Nations Development Programme, Global Dashboard for Vaccine Equity https://data.undp.org/vaccine-equity/ [https://perma.cc/P77Y-GJEN] (detailing the importance of data in achieving vaccine equity).

26. *Cf.* Nature Editorial, *Nature Addresses Helicopter Research and Ethics Dumping*, 606 NATURE 7 (2022); AFRICAN ACADEMY OF SCIENCES, RECOMMENDATIONS FOR DATA AND BIOSPECIMEN GOVERNANCE IN AFRICA (Feb. 8, 2021), https://openresearchafrica.s3.eu-west-1.amazonaws.com/posters/docs/aasopenres-184857.pdf [https://perma.cc/W8LT-KJ4Y] (detailing the concerning exclusion of African researchers from the IP and development stages).

27. *See* Stephanie Russo Carroll et al., *Extending the CARE Principles from Tribal Research Policies to Benefit Sharing in Genomic Research*, 13 FRONTIERS IN GENETICS (Nov. 11, 2022) (on gaps to address for substantive benefit sharing and a need to strengthen Responsibility and Ethics in tribal research governance).

28. Ilene Karsch-Mizrachi et al., *The International Nucleotide Sequence Database Collaboration,* 40 NUCLEIC ACIDS RESEARCH D33 (2012).

The making, sharing, and use of these sequences are mainly the work of biological scientists or informatics and data scientists around the world. In that respect, the production and flow of sequences has in the first instance been determined by scientist norms of cooperation and credit, overlain by some rules and policies of funders, journals, and sequence-data infrastructure providers. Channeling, overhanging, and at times blocking this are a diverse assortment of legal structures, disciplinary and institutional practices, power-political interests, and normative demands for (equal) rights and justice. The arguments in this article about attributive justice in GSD are largely applicable no matter what kind of sample is being sequenced: viruses, bacteria, plants, human and other animal life, and which of the 'omics' is involved (genomics, proteomics etc.). The sequencing techniques and GSD infrastructures too are broadly similar across this whole range of samples. Additional issues of governance, personal data protection, human rights, and for some countries even national security are implicated when human genomic material is involved—research focused on virus GSD seeks to cabin these issues by anonymizing and restricting access to human data and metadata associated with a virus sample.[30] Viruses as a category has the complication that vastly more viruses are undiscovered than are thus far known, and even for known viruses their effects are often not known.[31] The small (but expanding) subset of viruses that are known to be pathogens or potential pathogens for humans or other populations has been the subject of a further layer of governance, initially for material sam-

29. *See* Bruno Strasser, *The Experimenter's Museum: GenBank, Natural History, and the Moral Economics of Biomedicine,* 102 Isis 60 (2011) (stating that the GenBank itself was established in 1982).

30. Rudolf I. Amann et al., *Toward Unrestricted Use of Public Genomic Data: Publication Interests Should Not Limit Access to Public Data*, 363 Science 350, 351 (2019). *See also* Amal Matar et al., *A Proposal for an International Code of Conduct for Data Sharing in Genomics*, Developing World Bioethics (Oct. 22, 2022) (discussing some of the data ethics and law questions relating to human genome sequences—which may apply also to some cases where a human tissue sample is the source of a virus which is then sequenced).

31. *See* Elizabeth Pennisi, *New Dangers? Computers Uncover 100,000 Novel Viruses in Old Genetic Data*, Science Magazine (Jan. 26, 2022), https://www.science.org/content/article/new-dangers-computers-uncover-100-000-novel-viruses-old-genetic-data (stating how scientist have recently found 100,000 novel viruses but estimate trillion more undiscovered may still exist).

ples and more recently for GSD. Pathogen data in general, and data on pathogens implicated in a major epidemic all the more so, is subject to demands to prioritize sharing that do not apply to other GSD.[32] However, contestation over principles of GSD pathogen data collection and sharing, including demands for attributive and distributive justice, also intensify during a severe epidemic. This results in a cycling pattern in which the intensity of interests and political engagement during a pandemic make it difficult to reach a permanent agreement, but weakening political interest in governance of pathogenic GSD after an epidemic subsides means that in inter-pandemic periods, the issues fall back either to technocratic and infrastructural bodies (such as the WHO Secretariat or the INSDC), or to political bodies whose primary concerns are with larger topics (such as the Convention on Biodiversity Conference of the Parties).[33]

Major changes in the means of scientific knowledge production and shifts in views and practices about what receives (or should receive) recognition and professional status, have precipitated significant debate about the suitability of its infrastructures. Three such sets of changes are particularly germane to this article.

First are ongoing efforts to extend the valorization of scientific labor beyond the recognition and status accorded to theoreticians and experimentalists by publications in prestigious journals. Active coalitions seek to valorize data production, data sharing, and projects built collectively from shared data. Rachel Ankeny and Sabina Leonelli in 2015 articulated widely held hopes that postgenomic biology would foster "new forms of labor and ideals of community and openness that differ

32. *See generally* World Health Organization, WHO Global Genomics Surveillance Strategy for Pathogens with Pandemic and Epidemic Potential (2021) (viewing high risks from pathogens as both justifying and necessitating a distinctive approach to pathogen surveillance and data sharing).

33. Independent Panel for Pandemic Preparedness and Response, COVID-19: Make it the Last Pandemic (World Health Organization, May 2021). The Co-Chairs of the Panel were Ellen Johnson Sirleaf and Helen Clark.

considerably from the period that preceded the postgenomic era."[34]

Second is the increased focus on systematic under-recognition in the credit and reward system of women scientists and many historically excluded groups,[35] under-recognition of many forms of labor and of the people (often women and minorities) performing it in scientific and data processes,[36] and initiatives to "decolonize" research and institutions.[37]

Third is that pressure on the existing regime has been intensified by the separation between digital genetic information (DGI) and the physical material (or location) from which it was initially derived. "De-materialization" has enabled many aspects of modern biology, from digitization of traditional natural history museum collections, to bioinformatics, to synthetic biology in which adjusted or new organisms can be produced in labs or a virus could be produced in a new place based only

---

34. Rachel Ankeny & Sabina Leonelli, *Valuing Data in Postgenomic Biology: How Data Donation and Curation Practices Challenge the Scientific Publication System, in* POSTGENOMICS: PERSPECTIVES ON BIOLOGY AFTER THE GENOME 126, 127 (Sarah Richardson and Hallam Stevens eds., 2015). *See also* Barbara Prainsack & Sabina Leonelli, *Responsibility, in* SCIENCE AND THE POLITICS OF OPENNESS: HERE BE MONSTERS (Brigitte Nerlich, Sarah Hartley, Sujatha Raman & Alexander Smith eds., 2018) (lamenting, and aspiring to reverse, what they see as a turn away from organizations of scientists engaged in a collective political responsibility "to ensure that science contributes to making our societies more just and more dignified for everybody to live in", to a view of responsibility as defined by innumerable codes of ethics and connoting also "a duty for individual scientists—professional scientists or citizen scientists—to be useful to existing systems").

35. *See* Margaret W. Rossiter, *The Matthew Matilda Effect in Science*, 23 SOC. STUD. OF SCI. 325 (1993) (discussing how woman scientist and their achievements are often erased and eclipsed by male counterparts labeling the phenomenon "Matilda effect"); *see also* Travis A. Hoppe et al., *Topic Choice Contributes to the Lower Rate of NIH Awards to African-American/Black Scientists,* 5 SCI. ADVANCES 7238 (2019) (examining the lower rate of NIH funding received by African American Scientists).

36. *See* Clémence Pinel, Barbara Prainsack & Christopher McKevitt, *Caring for Data: Value Creation in a Data-Intensive Research Laboratory*, 50 SOC. STUD. OF SCI. 175 (2020) (discussing the role of care in data work and how it is undervalued and gendered).

37. Mishal Khan et al., *Decolonising Global Health in 2021: A Roadmap to Move from Rhetoric to Reform*, 6 BMJ GLOBAL HEALTH e005604 (2021); Lioba A. Hirsch, *Is it Possible to Decolonise Global Health Institutions?,* 397 LANCET 189 (2021).

on online access to the digital nucleotide sequence.[38] For regulatory purposes, the key consequence of de-materialization is that it calls for data governance, rather than physical materials governance, which many of the existing regulatory regimes have been premised on.[39] As digital sequences travel separately from biological samples and may combine material from multiple samples or become disconnected from the original source, states of origin struggle to exercise much or any functional jurisdiction over them. Moreover, customs and border controls as well as freight and transit rules that are applicable to physical samples for most states are not replicated in control over digital information flows.[40] The entire system of Materials Transfer Agreements, which has become widespread practice in governing scientific and commercial transfer of biological research samples, is not readily adapted to digital sequences distributed through public databases. Yet, global data governance structures are not developed enough to do the work currently done by physical materials regulation.[41] This article focuses on GSD that arises in some way from a biological sample.[42]

---

38. Margo Bagley, *De-Materializing Genetic Resources: Synthetic Biology, Intellectual Property and the ABS Bypass, in* ROUTLEDGE HANDBOOK OF BIODIVERSITY AND THE LAW 219 (Charles McManis and Burton Ong eds., 2017).

39. *See* Eric T. Juengst & Eric M. Meslin, *Sharing with Strangers: Governance Models for Borderless Genomic Research in a Territorial World*, 29 KENNEDY INSTITUTE OF ETHICS JOURNAL 67 (2019) (comparing four governance models for data-sharing policy and practices in genomics); *see also* Mahsa Shabani, *Blockchain-Based Platforms for Genomic Data Sharing: A De-Centralized Approach in response to the Governance Problems?*, 26 J. AM. MED. INFORM. ASSOC. 76 (2019) (discussing blockchain-based platforms as a solution to emerging technical and governance challenges involving genomic data sharing).

40. Charles Lawson et al., *The Future of Information Under the CBD, Nagoya Protocol, Plant Treaty, and PIP Framework*, 22 J. WORLD INTELL. PROP. 103 (2019)

41. *See generally* Michelle Rourke, *Origin of Samples: Pathogen Provenance and the Rise of the Material Transfer Agreement*, 3(2) JOURNAL OF SCIENCE AND LAW 1–3 (2017) (explaining the inadequacies of current data governance structures under the United Nations Convention on Biological Diversity and domestic access and benefit-sharing laws).

42. It does not consider the somewhat different problems arising in the production *in silico* of purely computer-generated sequences, nor does it address the numerous issues arising from fast-moving developments in synthetic biology. *See, e.g.*, Hung-En Lai et al., *Synthetic Biology and the United Nations*, 37 TRENDS IN BIOTECHNOLOGY 1146 (2019) (mentioning numerous issues arising from fast-moving developments in synthetic biology); *see also*

Section II provides a brief explanation of the nature and significance of virus genomic sequence data sharing, the infrastructures through which it takes place transnationally, and some factors driving the extent and forms of this data sharing. Section III examines scientist cultural-professional norms which might be subsumed into ideas of attributive justice: attribution of work, first-dibs opportunities in publication, and ancillary normative practices with regard to such matters as peer-review and data sharing. Section IV addresses attributive justice claims made by states in relevant national and international law, invoking ideas such as virus sovereignty, exchanges of access to biological-materials in return for benefit sharing, and data sharing. It notes the mismatches between an exclusive attributive justice focus and the kinds of universal arguments made in human rights claims to basic healthcare and to participate in the benefits of science. Section V points to the increasing importance of infrastructure and of infrastructural regulation in shaping the possibilities of attributive justice in relation to pathogen sequence data sharing, and the normative and practical value of enhancing publicness in data-sharing infrastructure.

II. Virus Genomic Sequence Data Sharing: What, How, Why

What is meant by a genomic sequence, beyond the first-cut answer that it is a series of nucleotides described by ordered configurations of standardized symbols used to designate particular physical substances (e.g., A, C, G, T for DNA; A, C, G, U for RNA)? A Science and Technology Studies (STS) approach might extend much further to include the laboratories and procedures, the knowledge conventions, etc. In between are pragmatic approaches driven by policy and political concerns. For example, in the deliberations within a technical group of the Convention on Biological Diversity, the following are asserted as among the "types of information that may be relevant to the utilization of genetic resources" in some cases:

Secretariat of the Convention on Biological Diversity, Synthetic Biology, CBD Technical Series No. 100 (2022), https://www.cbd.int/doc/publications/cbd-ts-100-en.pdf [https://perma.cc/2JAH-96LA] (discussing in detail issues with synthetic biology governance).

> (a) The nucleic acid sequence reads and the associated data;
> (b) Information on the sequence assembly, its annotation and genetic mapping. This information may describe whole genomes, individual genes or fragments thereof, barcodes, organelle genomes or single nucleotide polymorphisms;
> (c) Information on gene expression;
> (d) Data on macromolecules and cellular metabolites;
> (e) Information on ecological relationships, and abiotic factors of the environment;
> (f) Function, such as behavioral data;
> (g) Structure, including morphological data and phenotype;
> (h) Information related to taxonomy;
> (i) Modalities of use.[43]

Debates over which of these types of information should actually be within the various conditioned data sharing and access-benefit regimes are in several cases highly contentious.

Sharing of genomic sequence information for epidemic viruses has the obvious predicates that in a specific situation there are available the skills and means to isolate the virus (or enough fragments of it) and conduct the sequencing, and that there exists an adequate method to circulate the GSD so that many other researchers have easy access to it in a format they can use.[44] In many situations these two pre-conditions—capacity for collection and sequencing, and an infrastructure for GSD circulation and readout—are not met. The COVID-19 epidemic renewed attention in the United States, as elsewhere, on insufficiency in the availability and use of virus genomic sequence data and to patchy, under-funded, and under-integrated data sharing structures. Calls were made for the federal government to "develop and invest in a national data infrastructure system that can link genomic, clinical, and epidemio-

---

43. Report of the Ad Hoc Technical Expert Group on Digital Sequence Information on Genetic Resources, Annex, para 2. UN Doc CBD/DSI/AHTEG/2018/1/4 (Feb. 20, 2018, reissued May 23, 2018).

44. *See* WHO, GLOBAL GENOMIC SURVEILLANCE STRATEGY FOR PATHOGENS WITH PANDEMIC AND EPIDEMIC POTENTIAL, 2022–2032, at 7, 8, 13 (Mar. 2022) (outlining objectives including strengthening technical workforces and enhancing data sharing systems).

logical data."[45] Many of the problems are legal, regulatory, and institutional—for example, a U.S. National Academies scientific panel in 2020 lamented that in the United States had "no current federal or state laws protect or mandate the sharing of sequence data from virus samples."[46]

Similar or greater infrastructural limitations have beset systematic large-scale virus genomic sequencing in many other countries and regions.[47] Virus surveillance and collection capacity is quite modest or even non-existent in some developing countries—acquisition of genetic sequencing machines and operating expertise has been low in budget priorities for the governments of most low-income countries, with health expenditures ranging from primary care clinics to child immunization and vaccination cold-chain storage understandably ranking much higher in urgency.[48] Development aid agencies, research institution collaborations, and specialized intergov-

---

45. National Academies of Science, Engineering, and Medicine, Consensus Study Report, Genomic Epidemiology Data Infrastructure Needs for SARS-CoV-2: Modernizing Pandemic Response Strategies (Jul. 2020) 'Highlights' Summary, 1–2.

46. *Id.*

47. Mary Aigbiremo Oboh et al., *Translation of Genomic Epidemiology of Infectious Pathogens: Enhancing African Genomics Hubs for Outbreaks*, 99 Int'l J. Infectious Diseases 449 (2020). "The current state of institutions, infrastructure and human resources for [African GSD] data generation, management and analysis needs urgent attention. To respond to epidemics such as COVID-19, genomic data generation for real-time decision making could be enhanced by the adoption and decentralised application of small, portable and easily operated experimental tools such as the Oxford Nanopore technology-MinION sequencer, Illumina MiniSeq or the BGI-DNBSeq across all countries. These easily deployable, user-friendly field-based technologies were instrumental in the sequencing of the Ebola virus (EBOV) during the last outbreak in West and Central Africa . . .The data generated were useful for strategizing and ensuring the efficacy of interventions, including tracking and stopping the spread of EBOV and evaluating vaccine efficacy." *Id.*, 451 (references omitted).

48. *See generally* Mohamed Helmy, Mohamed Awad, & Kareem A. Moosa, *Limited Resources of Genome Sequencing in Developing Countries: Challenges and solutions*, 9 Applied & Translational Genomics 15 (2016) (explaining that limited R&D funding and other budget-related issues affect the availability of sequencing technologies in developing countries); *see also* Abdoulie Kanteh et al., *Simple and Structured Model to Build Sequencing Capacity in West Africa*, 10 The Lancet Glob. Health e1240 (2022) (noting the scarcity of genetic sequencing infrastructure and skills as a hinderance in the management of the COVID-19 outbreak in Africa).

ernmental organizations have supported major public health and research programs with some virus genomic sequencing capacity,[49] including the creation of the Africa Centres for Disease Control and Prevention in Addis Ababa.[50] One effect of the COVID-19 pandemic was ramped up support (including donations from some major sequencing machine manufacturers) for sequencing hubs in many developing countries,[51] and prioritization of local sequencing capacity over reliance on sending samples abroad.[52] However, capabilities continue to be highly uneven, and sustaining local hubs and expertise at inter-pandemic times of low demand remains challenging.

These disparities and limitations arise against the overall context of a significant infrastructure for transnational virus GSD sharing coming into operation and expanding rapidly

---

49. *See* Greg Cima, *Pandemic Prevention Program Ending After 10 years: USAID Predict Led Virus Discovery, Health Training, Risk Education*, JAVMA NEWS (Jan. 15, 2020) (outlining the successes of the PREDICT program). *See also* Luke Lythgoe, *Preventing Future Zoonotic Pandemics*, GLOBAL COMMISSION FOR POST-PANDEMIC POLICY, POLICY MONITOR (April 5, 2021) https://globalcommissionforpostpandemicpolicy.org/preventing-future-zoonotic-pandemics/ [https://perma.cc/6YT6-X986] (describing some such programs to prevent zoonotic pandemic); *but see* Ann Linden & Dale Jamieson, *Blind Spots in Biodefense*, 379 SCI. 621, 621 (2023) (pointing to continuing deficiencies in zoonotic disease surveillance within the United States).

50. John N. Nkengasong et al., *Establishing the Africa Centres for Disease Control and Prevention: Responding to Africa's Health Threats*, 5 THE LANCET GLOBAL HEALTH e246 (2017).

51. *See* WHO Africa, *Scaling Up Genomic Sequencing in Africa* (Sept. 30 2021), https://www.afro.who.int/news/scaling-genomic-sequencing-africa [https://perma.cc/T72S-UXZM] (stating that "[i]n 2020, WHO and the Africa Centre for Disease Control and Prevention established a COVID-19 sequencing laboratory network in Africa which has to date produced over 43 000 sequencing data."); *see also* WHO Africa, *Scaling Up Genomic Sequencing in Nigeria* (Feb. 14, 2022), https://www.afro.who.int/photo-story/scaling-genomic-sequencing-nigeria [https://perma.cc/2JLW-TAHE] (stating that "[t]he Organization has also provided polymerase chain reaction (PCR) screening kits for rapid detection of variants of concern, laboratory consumables for PCR and sequencing as well as sample transport and storage equipment to Nigeria and other countries in Africa.").

52. *See generally* VIJAY SHANKAR BALAKRISHNAN, WHO, *Global Genomic Surveillance Strategy for Pathogens with Pandemic and Epidemic Potential, 2022–2032* (2022), https://www.who.int/publications/i/item/9789240046979 [https://perma.cc/D5G6-G7X5] (providing background on local sequencing) (providing background on local sequencing).

from the 1980s onward.[53] Enormous development of computing power and software, as well as high-capacity data transmission and cloud storage, have facilitated the growth of very large databases and informatics. Two major genomic sequence databank infrastructures used transnationally are the set grouped in the INSDC—discussed here alongside GenBank, its U.S. member body—and GISAID-EpiFlu.

GenBank was founded in 1982 and is based and administered within the U.S. National Institutes of Health (NIH) by the National Center for Biotechnology Information (NCBI),[54] which also operates several associated databases, such as the Reference Sequence (RefSeq) database.[55] The direct precursor to GenBank was the Los Alamos Sequence Library, a collection of sequences made by staff at the Los Alamos National Laboratory (LANL) from the late-1970s, and LANL was the first host of GenBank.[56] GenBank is part of the International Nucleotide Sequence Database Collaboration (INSDC), a data sharing consortium of databanks providing similar services with worldwide access to anyone, in which the DNA Data Bank of Japan and the European Molecular Biology Laboratory's European Bioinformatics Institute are also founding members.[57] By 2019, the INSDC (and hence GenBank) held over

---

53. *See generally* Strasser, *supra* note 29 (describing the establishment of the GenBank); *See also* HALLUM STEVENS, LIFE OUT OF SEQUENCE: A DATA-DRIVEN HISTORY OF BIOINFORMATICS (2013) (chronicling the emergence of bioinformatics, starting from the 1960s).

54. Strasser, *supra* note 29, at 60 (2011).

55. *See* Arita et al., *supra* note 4, at D121 (2021) (explaining that the NCBI operates several databases that link to the GenBank raw sequence database but may accommodate other materials, or materials with restrictions (such as human genomic materials subject to privacy-based or consent-based restrictions).

56. *See* Strasser, *supra* note 29 (explaining that the major competing bid to design and host GenBank in 1982 came from Margaret Dayhoff and the National Biomedical Research Foundation, a private non-profit institution near Washington DC, and also noting that Margaret Dayhoff had initiated the *Atlas of Protein Sequence and Structure* in 1965 and was a leading figure in sequence curation and re-publication).

57. Hallam Stevens, *Globalizing Genomics: The Origins of the International Nucleotide Sequence Database Collaboration,* 51 J. HIST. OF BIOLOGY 657, 658 (2018). The INSDC synchs the sequences held by the three databases. In 2019 the cost of running the INSDC was between USD 50-60 million per year. *Combined Study on Digital Sequence Information in Public and Private Databases and Traceability*, CBD/DSI/AHTEG/2020/1/4 at 3 (Jan. 31, 2020).

212 million entries; most quite short (85% are under 1000 bases), but some were six orders of magnitude longer. About twelve percent were human genomic sequences, and another twelve percent were model organisms (e.g. fruit fly), while the remainder were (in descending order of number of sequences) animals, plants, bacteria and micro-organisms, fungi, and viruses.[58]

The GISAID Initiative (formerly the Global Initiative on Sharing All [previously Avian] Influenza Data) began operations in 2008 and runs the EpiFlu database.[59] GISAID is administered by a purpose-designed German non-profit foundation. It has received support from the governments of Germany (which provided GIASID's platform hosting from 2010–21), Singapore, and the United States, along with other entities.[60] Anyone can join GISAID and submit or download virus sequences, but the person creating an account must provide proof of identity, and use of data from GISAID is governed by conditions including making efforts to work jointly on publication with the originating lab, and attribution of credit to the original producers of the sequence.[61] GISAID was founded in support of research on influenza viruses, but from January 2020 became a major databank for COVID-19 research. GISAID was the entity to which the full sequences of the SARS-CoV-2 virus were submitted by Chinese Centers for

58. *Id.* at 28.

59. Stefan Elbe & Gemma Buckland-Merrett, *Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health*, 1 GLOB CHALL. 33 (2017).

60. *History*, GISAID, https://gisaid.org/about-us/history/ [https://perma.cc/JSS2-UM8Q] (last visited Apr. 17, 2023). Significant criticisms of leadership style and communications, uneven access and erratic denials of access to sequences, and overall governance arrangements of GISAID were crystallized in a report in *Science* in April 2023. Martin Enserink and Jon Cohen, *Control Issues*, 380 SCIENCE 332 (2023). This article reported significant concerns among leading science research funders and some scientists, largely in Europe and North America (*id.* at 337-9), but also strong appreciation from some scientists in developing countries (*id.* at 337). GISAID faces considerable pressure for governance improvement if it is to continue to flourish, but it also has considerable support. This article is conceptual and does not examine these important institutional matters with regard to GISAID (or INSDC).

61. *Frequently Asked Questions*, GISAID, https://gisaid.org/help/faq/ [https://perma.cc/9H6X-SDZJ] (last visited Apr. 17, 2023).

Disease Control researchers and distributed on January 12, 2020.[62]

For sequence viewing and downloading, many different infrastructural designs of access to online virus GSD databanks could be established, each resulting in somewhat different practical arrangements and ethical or contractual conditions or other legal obligations for users, as well as variance in the rights of original resource-holders or sequencers or others and in the practical enforceability of thee rights.[63] These conditions may be imposed on sequence submitters or those introducing annotations or links or metadata, sequence viewers and downloaders, and those wishing to use data analytics and other compute functions on the platform. These complexities are the embodied details which are fundamental to the real operation of infrastructures. As the present article is concerned only with conceptual considerations, it simply represents the proclaimed differences among key GSD infrastructures through a highly reductionist three-fold stylization. A *private-access* system restricts access to sequences, by requiring advance permission to read or download GSD, granted only to those paying a subscription, or to those receiving security clearance, etc.; this may be enforced technologically by requiring a centrally supplied password or other credentialing tokens. Upload and compute are usually restricted too. Private biological research companies often hold GSD in this way. A *conditioned open-access* system allows anyone to have access to the databank, but access

62. *See China Delayed Releasing Coronavirus Info, Frustrating WHO* Associated Press, June 4, 2020, https://apnews.com/article/united-nations-health-ap-top-news-virus-outbreak-public-health-3c061794970661042b18d5aeaaed9fae [https://perma.cc/2H96-F7K9] (explaining that GISAID, notably, had until then been only for influenza virus sequences, not betacoronaviruses and that other whole- or part-genome sequences of this virus were submitted to other sequence sharing infrastructures abroad by other scientists from China slightly before the release by the Chinese CDC, and perhaps precipitated the CDC to act); *see also* Fan Wu et al., *A New Coronavirus Associated with Human Respiratory Disease in China*, 579 Nature 265, 265–269 (2020) (explaining that SARS-CoV-2 is a positive-sense, single-strand RNA virus and that the whole-genome NCBI Reference Sequence of SARS-CoV-2, isolated from a patient in Wuhan in late 2019, is 29,903 nucleotides in length).

63. Marcel Jaspars & Abbe E. L. Brown, *What Should We Mean by 'Open Access'?, in* Access and Benefit Sharing of Genetic Resources, Information and Traditional Knowledge 89 (Charles Lawson, Michelle Rourke, & Fran Humphries eds., 2023).

to the databank is controlled by a log-in and password or other credentialing system, and made subject to the user agreeing in advance to certain conditions of use.[64] Virtually all digital infrastructures, including those of the INSDC, impose some conditions for platform access and use (on the submit side and the use side), and hence are regulatory.[65] GISAID's conditions have become prominent not because such infrastructural regulation is unusual, but because they restrict access as a means for influencing external norms and behavior in a direction different from the way in which openness has been defined in the INSDC model. At least when they become well-known and can readily be contrasted with conditions set by other infrastructures, the sets of conditions are in themselves normative, expressive, and performative. However, it is not only the conditions but also the means and practices of enforcement that are potentially significant to social understandings. Enforcement against those who access the databank but then violate the conditions is potentially undertaken by excluding them from further use of the databank, by reputational sanctions, and possibly through private civil actions using state law systems.[66] This seems to work adequately amongst connected professional science communities but is likely insufficient to deter a

64. See for example the NATIONAL HUMAN GENOME RESEARCH INSTITUTE, https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science [https://perma.cc/63NH-MKKY] (explaining that the U.S. National Human Genome Research Institute uses the terms 'registered access' (access requires prior registration, and the user may be monitored) and 'controlled access' (access is granted only after approval))

65. For example, the Database of Genotypes and Phenotypes (dbGaP), operated (as GenBank is) under the auspices of the U.S. National Library of Medicine. In Contreras's summary: "dbGaP can accommodate phenotypic data, which includes elements such as de-identified subject age, ethnicity, weight, demographics, drug exposure, disease state, and behavioral factors, as well as study documentation and statistical results. Given potential privacy and regulatory concerns regarding phenotypic data, dbGaP allows access to data on two levels: open and controlled. Open data access is available to the general public via the Internet and includes non-sensitive summary data, generally in aggregated form. Data from the controlled portion of the database may be accessed only under conditions specified by the data supplier, often requiring certification of the user's identity and research purpose." Jorge Contreras, *Optimizing Access Policies for Big Data Repositories: Latency Variables and the Genome Commons*, *in* BIG DATA OPTIMIZATION: RECENT DEVELOPMENTS AND CHALLENGES 201, 204 (Ali Emrouzejnad ed., 2016).

66. WORLD HEALTH ORGANIZATION, *supra* note 14, at 5.

highly motivated violator who wishes to challenge the arrangements for ideational or symbolic reasons, or has a major material interest to advance through breach. An *unconditional open-access* system allows anyone to use data from the databank for any otherwise-lawful purpose, including research, publication, and commercial development.[67] This is how the INSDC/GenBank model is portrayed on the user-access side, although this model is of course regulatory and expressive in its own way.[68] The INSDC argues that the term 'open access' does not adequately summarize its approach, because access may be open but use is restricted e.g. by the terms of a license, whereas the INSDC members such as GenBank for the most part do not allow such restrictions, and any constraints that may arise from intellectual property rights or other laws and

67. Slightly more restricted versions sometimes use the term 'open access'. *See e.g.*, *Genomic Data Science Fact Sheet*, NAT'L HUM. GENOME RSCH. INST. (Apr. 6, 2023), https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science [https://perma.cc/4Q3Y-7482] (stating that the data is available for research purpose, leaving out purposes such as commercial development and publication).

68. Each of these models may also involve the databank setting some conditions on upload of material to the databank, such as the condition that the uploader has full authority to do so, and the condition that no restrictions on access or use can be imposed by the sequencing or the sample-originating laboratory other than those set by the databank itself. Such conditions may demand that the uploader grant general permission to use the GSD, similar to an open-source software license. *See generally* Masanori Arita, *Open Access and Data Sharing of Nucleotide Sequence Data*, *in* MULTIDISCIPLINARY DATA ACTIVITIES BRIDGING THE RESEARCH COMMUNITY AND SOCIETY (Tomoya Baba et al. eds., 2021) (written while the author was head of the DDBJ, one of the core members of INSDC). The INSDC proclaims it "has a uniform policy of free and unrestricted access to all of the data records their databases contain. Scientists worldwide can access these records to plan experiments or publish any analysis or critique. Appropriate credit is given by citing the original submission, following the practices of scientists utilizing published scientific literature.

> The INSDC will not attach statements to records that restrict access to the data, limit the use of the information in these records, or prohibit certain types of publications based on these records. Specifically, no use restrictions or licensing requirements will be included in any sequence data records, and no restrictions or licensing fees will be placed on the redistribution or use of the database by any party. All database records submitted to the INSDC will remain permanently accessible as part of the scientific record." (numbering omitted) *Policy*, INSDC, https://www.insdc.org/policy/ [https://perma.cc/M2A9-Q9JZ] (last visited on May 10, 2023).

contracts are implemented outside the infrastructure and not inside it.[69]

A well-established approach in scholarly investigations of infrastructure emphasizes the interacting technical, organizational, and social aspects that fundamentally characterize nearly all infrastructure. "For our purposes, cyberinfrastructure is the set of organizational practices, technical infrastructure and social norms that collectively provide for the smooth operation of scientific work at a distance. All three are objects of design and engineering; a cyberinfrastructure will fail if anyone is ignored."[70] This analytic structure readily applies to the GSD infrastructures used for pathogenic virus sequences and metadata.[71]

A. *Technical*

Both GenBank and GISAID are electronic databanks, storing nucleotide sequence information as well as metadata relating to the original biological sample and to the sequencing. They include full genome or exome (the protein-encoding parts of the genome) nucleotide sequences of viruses and other pathogens isolated from humans and many other organisms, as well as protein structures and other sequences.[72] From early 2020 onward, they each held large and fast-increasing collections of SARS-CoV-2 virus sequences.[73] These collec-

---

69. Arita et al., *supra* note 4, at D122.

70. PAUL EDWARDS, STEVEN JACKSON, GEOFFREY BOWKER & CORY KNOBEL, UNDERSTANDING INFRASTRUCTURE: DYNAMICS, TENSIONS, AND DESIGN—REPORT OF A WORKSHOP ON HISTORY AND THEORY OF INFRASTRUCTURE: LESSONS FOR NEW SCIENTIFIC CYBERINFRASTRUCTURES 6 (Jan. 2007). The same definition and accompanying text are in Geoffrey Bowker, Karen Baker, Florence Millerand & David Ribes, *Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment, in* INTERNATIONAL HANDBOOK OF INTERNET RESEARCH 97, 102 (Jeremy Hunsinger, Lisbeth Klastrup & Matthew Allen eds., 2007).

71. Florence Millerand & Geoffrey Bowker, *Metadata Standards: Trajectories and Enactments in the Life of an Ontology, in* STANDARDS AND THEIR STORIES: HOW QUANTIFYING, CLASSIFYING, AND FORMALIZING PRACTICES SHAPE EVERYDAY LIFE 149 (Martha Lampland & Susan L. Star eds., 2009).

72. On sequencing platforms see Jonathan Foox et al., *Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study*, 39 NAT. BIOTECH. 1129, 1129 (2021) (providing a relevant example).

73. Kirill Kryukov, Lihua Jin, & So Nakagawa, *Efficient Compression of SARS-CoV-2 Genome Data Using Nucleotide Archival Format*, 3:9 PATTERNS 100562 (2022).

tions, and the computing and informatics capacity deployed in relation to them, are of immense importance to biological research.[74] Participating laboratories and individual scientists contribute sequences by uploading them electronically. The databanks examine the submissions for anomalies, errors, and duplications before posting, and issue a unique accession number which can then be utilized in all publications or other materials referring to this sequence, enabling the GSD to link to related outputs in various forms. The personnel or algorithms of the databank may also check on necessary or desirable metadata (including details of the originating and sequencing scientists and their laboratories).[75] These curatorial services are increasingly performed by the databank's software, but human review and correspondence with submitters is also part of the process.[76] The infrastructures provide software and links to related databases, enabling sequence assembly, annotation, and bioinformatics research. Many important databanks are now cloud-based, usually in cooperation with commercial cloud services. Some provide online facilities for researchers to run software and store bioinformatics results, all within the databank servers—bringing their research queries and software to the data rather than vice versa.[77]

---

74. In relation to viruses with pandemic potential, GSD (including metadata) is fundamental to understanding the nature and pathogenic features of the virus, its origins and mutations (phylogenetics and clades), epidemiological information on its spread and origins (including zoonotic crossovers), potential for development of therapies and vaccines, as well as antibodies and antigens. *See, e.g.*, *Influenza Virus Genome Sequencing and Genetic Characterization*, CTR. FOR DISEASE CONTROL AND PREVENTION (Apr. 9, 2023), https://www.cdc.gov/flu/about/professionals/genetic-characterization.htm [https://perma.cc/82U6-ZDXS] (for an example in the United States).

75. *See, e.g.*, Tahani Nadim, *Data Labours: How the Sequence Databases GenBank and EMBL-Bank Make Data*, 25 SCI. AS CULTURE 496, 497 (2016) (providing examples such as BGI).

76. *See* SABINA LEONELLI, DATA-CENTRIC BIOLOGY: A PHILOSOPHICAL STUDY (Univ. of Chicago Press 2016) (explaining that some aspects of the culture of digital databank curation derive from that developed for biological samples in natural history museums and other collections). *See also* Ane M. Gabrielsen, *Openness and trust in data-intensive science: the case of biocuration*, 23 MED., HEALTH CARE AND PHIL. 497, 498 (2020) (emphasizing that the trust in science depends on the trust in the peers reviewing the work product before publication).

77. *See, e.g.*, Nadim, *supra* note 75, at 497 (providing GenBank as an example).

Some of these databanks are also important in establishing and designating the "reference" genomic sequence of each organism or pathogen.[78] Establishing the reference sequence for a novel virus with pandemic potential is important to enabling reliable genomic recognition of new suspected cases. However, mutations, recombinations, or other variations, including cross-overs with viruses sourced from other species, make imperative a large databank of sequences. Variation is thus one driver of databank scale. The speed of mutation and of disease contagion influences the rate at which it is necessary to generate and upload new sequences.[79]

Finding the virus genome material in a biological sample is not an easy process.[80] A great deal of labor and expertise is required in the labs, as well as the great effort and courage often involved in collecting the sample in the field.

B. *Organizational*

Like other large science infrastructures, GenBank-INSDC and GISAID each depend on public or science-focused philanthropic funding. Their data storage and access as well as curatorial services are supplied without charge to researchers. Were these costs not covered by public funding or philanthropy, other business models would have to be developed. The barriers to entry faced by new suppliers of infrastructure are quite high unless a profitable company saw a business case for investing in such infrastructure and could persuade users to upload sequences.

The INSDC institutions are each nested in some way in publicly funded science institutional structures. These institutions make major efforts to remain neutral to different companies and different university research groups. But it is ex-

78. *See, e.g.*, *RefSeq: NCBI Reference Sequence Database*, NAT'L CTR FOR BIOTECH. INFO. (April 9, 2023), https://www.ncbi.nlm.nih.gov/refseq [https://perma.cc/2GCW-62RH] (for U.S. NBCI's reference sequence database).

79. Madeleine Oman, Aqsa Alam, Rob W. Ness, *How Sequence Context-Dependent Mutability Drives Mutation Rate Variation in the Genome*, 14:3 GENOME BIOLOGY AND EVOLUTION evac032 (2022).

80. *See, e.g.*, *RNA-seq Sample Guidelines*, PENN STATE HUCK INSTITUTES OF THE LIFE SCIENCES, https://www.huck.psu.edu/core-facilities/genomics-core-facility/sample-recommendations/rna-seq-sample-guidelines [https://perma.cc/J5WC-KNZ9] (last visited May 10, 2023) (outlining the challenges of RNA sequencing).

tremely difficult for them to remain neutral in global terms when they are located within particular national (or EU) funding systems and subject to some level of political accountability. This was evident, for example, in pressure applied in 2020–2021 on the NIH by U.S. politicians, concerned about National Institute of Allergies and Infectious Diseases (NIAID) funding of gain-of-function virus research conducted in part by the Wuhan Institute for Virology.[81] This particular instance provided geopolitical overlay on a wider and longstanding debate about the scientific value and risks of funding or conducting this kind of research whether in the United States or elsewhere.[82]

GISAID was briefly connected to WHO, but quickly evolved into a German nonprofit entity which has received financial support from some governments and from major private foundations.[83]

C. *Social*

The reputation-authorship-credit moral economy of science combines historical traditions and inheritances with specificities of place and time. Bruno Strasser makes this case in positioning GenBank as a hybrid, fusing two important historical scientific traditions.[84] One is the building of unified collections of biological or other scientific specimens which are used to classify and compare within a broad field.[85] The other is the experimentalist sensibility that major advances in science can be generated in the lab through well-designed experiments and analysis of results.[86] GenBank is a massive collection, not of samples but of digitally rendered sequences of nucleotides, with accompanying metadata and phenotype information. The "collector" (NCBI) does not enjoy the controlling and pri-

---

81. Max Kozlov, *Risky 'Gain-Of-Function' Studies Need Stricter Guidance, Say US Researchers*, 605 NATURE 203, 204 (2022).

82. Updates to the U.S. government policy are noted on the NIH webpage, *Research Involving Enhanced Potential Pandemic Pathogens*, NATIONAL INSTITUTES OF HEALTH, https://www.nih.gov/news-events/research-involving-potential-pandemic-pathogens (last visited May 10, 2023).

83. Enserink & Cohen, *supra* note 60, at 339.

84. BRUNO STRASSER, COLLECTING EXPERIMENTS: MAKING BIG DATA BIOLOGY (Univ. of Chicago Press 2019) at 16, 224–5.

85. *Id.* at 224–5.

86. *Id.*

mary-use rights to the collection that had been the accepted standard for most collections of samples, but it does curate and organize the collection.[87] The collection is open to everyone to use, in the experimentalist-biology tradition of open science.[88] Galison and Daston emphasize historicized social relations in their historical study of natural science research collectivities, arguing that they "borrow and bricolage forms of labor in their ambient societies," and channel other forms of authority in the society (seeking royal patronage or aristocratic members, for example), but also tend to "invent new ways of working, communicating, adjudicating, and authorizing . . . because their aims are, in contrast to ordinary polities, epistemological, as well as economic, social, and political."[89]

What prompts sharing of virus genomic sequence data? A "commons" has long been formed by broadly available published research, often supported by publication or open-access deposit of underlying data.[90] The ethos of unhindered scientific collaboration and results sharing is very strong in research science communities across much of the world and is quite frequently mobilized to push back against governmental or commercial counterforces. Research scientists also have a strong interest in both the production and accessibility of a commons of underlying or associated data. The circulation of unpublished scientific research data, however, involves not only norms and practices of different scientific communities, but also commercial interests, policies of government agencies and, in some cases, international organizations and geopolitical or security concerns touching high-level political interests of states.

Probably the most prevalent invocation in support of infrastructures like GenBank-INSDC is "Open Science." This slo-

87. *See* Nadim, *supra* note 75, at 498 (explaining the roles as collectors played by GenBank and EMBL-Bank).

88. STRASSER, *supra* note 84.

89. Peter Galison & Lorraine Daston, *Scientific Coordination as Ethos and Epistemology, in* INSTRUMENTS IN ART AND SCIENCE: ON THE ARCHITECTONICS OF CULTURAL BOUNDARIES 296 at 297–98 (footnotes omitted) (Helmar Schramm, Ludger Schwarte & Jan Lazardzig eds., 2008).

90. GOVERNING KNOWLEDGE COMMONS 44 (Brett M. Frischmann, Michael J. Madison, & Katherine J. Strandburg eds., 2014); Jorge Contreras & Bartha Knoppers, *The Genomic Commons,* 19 ANNUAL REVIEW OF GENOMICS AND HUMAN GENETICS 429, 437 (2019).

gan aligns practices prevailing among (some) scientists and their professional collectivities with one set of broader politico-institutional policies and understandings in relation to science.[91] "Open Science" in both "Whiggish" and "naturalist" accounts of the post-1945 West reflected a set of changed circumstances pithily summarized by Shapin and Schaffer as "the ever-more-intimate enfolding of scientific research in the institutions of the state," and "the cultural celebration of science as integral to national security and economic welfare."[92]

A much less high-minded approach to "Open Science" is now also articulated in North Atlantic polities, eschewing "ethos" and advocating instrumental and pragmatic calculus. For example, the 2020 report of a group advising the European Commission asserted: "Open Science for its own sake has never been the goal. Openness is a vital instrument which, when used responsibly, can fuel a faster, more effective, more reliable, more trustworthy, more equitable and more innovative shared research knowledge system."[93]

---

91. Section 6 of UNESCO Recommendations on open science defines "open science" as combinations of "various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community." Section 13 on Core Values proclaims: "b. Collective benefit: as a global public good, open science should belong to humanity in common and benefit humanity as a whole. To this end, scientific knowledge should be openly available and its benefits universally shared. . . c. Equity and fairness: open science should play a significant role in ensuring equity among researchers from developed and developing countries, enabling fair and reciprocal sharing of scientific inputs and outputs and equal access to scientific knowledge." With some irony, the UNESCO graphic designers included on this page of the formal document (catalog number 0000379949, at p. 17) a compass graphic with a bold arrow pointing North and the rest all shaded. UNESCO Recommendation on Open Science (November 2021), available at https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=EN [https://perma.cc/J2LY-X7TE].

92. Steven Shapin & Simon Schaffer, *Introduction to the 2011 Edition: Up for Air—Leviathan and the Air-Pump a Generation On, in* LEVIATHAN AND THE AIR-PUMP: HOBBES, BOYLE, AND THE EXPERIMENTAL LIFE xxii (Steven Shapin & Simon Schaffer eds., 2nd ed. 2011).

93. *See* EU OPEN SCIENCE POLICY PLATFORM, PROGRESS ON OPEN SCIENCE: TOWARDS A SHARED RESEARCH KNOWLEDGE SYSTEM—FINAL REPORT OF THE OPEN SCIENCE POLICY PLATFORM at 22, 26 (European Commission 2020).

The interest of scientists and of information-intensive innovation-focused economies in "Open Science" is reinforced in the case of pathogen data by its obvious value to health, the economy, and national security. The mobilization of these assorted interests is exemplified by the Global Health Security Initiative (GHSI), a George W. Bush-era coalition of the willing established in 2001 among Canada, the European Commission, France, Germany, Italy, Japan, Mexico, the United Kingdom, and the United States—the World Health Organization was designated an expert advisor.[94] After GHSI members struggled to obtain MERS-CoV samples at a time when MERS seemed a potential pandemic threat, they created the GHSI Sample Sharing Task Group (SSTG), and committed to "support open, transparent, and rapid sharing [of samples] to facilitate a timely public health response" and to share pandemic-potential samples with each other and the WHO.[95]

In evaluating the interacting technical, organizational, and social dimensions of infrastructures such as GISAID and INSDC, it is necessary also to consider the relations of these to

---

This report bluntly stresses competitive concerns: "The right balance between Open Science, the potential to maximize the use and re-use of research data and outputs, IPR, and private companies' competitiveness must be promoted and become a central feature of the next round of discussions on the future of a shared research knowledge system. There are limits to openness and these must be acknowledged and taken into account as the system changes . . . Dissemination of research knowledge should also take place on a reciprocal basis, especially at an international level . . . Open Science policies can boost the performance of both the European economy and global economy, while IPR ensures the added value falls within European boundaries when appropriate." A similar idea was articulated in more theoretical terms by Paul David, *The Economic Logic of "Open Science" and the Balance between Private Property Rights and the Public Domain in Scientific Data and Information: A Primer*, *in* THE ROLE OF SCIENTIFIC AND TECHNICAL DATA AND INFORMATION IN THE PUBLIC DOMAIN: PROCEEDINGS OF A SYMPOSIUM 19, 22 (National Academies ed., 2003).

94. Maria Julia Marinissen et al., *Sharing of Biological Samples During Public Health Emergencies, in* VIRAL SOVEREIGNTY AND TECHNOLOGY TRANSFER: THE CHANGING GLOBAL SYSTEM FOR SHARING PATHOGENS FOR PUBLIC HEALTH RESEARCH 155, 167 (Sam Halabi & Rebecca Katz eds., 2020). On coalitions of the willing, see generally Alejandro Rodiles, *Coalitions of the Willing in Context: The Interplay between Formality and Informality, in* COALITIONS OF THE WILLING AND INTERNATIONAL LAW 148 (Cambrigde Univ. Press 2018).

95. Marinissen et al., *supra* note 94.

other norms and practices with which they interact.[96] For example, by setting requirements that authors of papers make publicly available the sequence data they have used, and rules about timing and format, a leading journal such as *Nature*[97] or an organization such as the International Committee of Medical Journal Editors[98] can generate almost inexorable incentives for authors, which in turn affect the infrastructures and public health entities seeking earlier and more comprehensive information on diseases. Funders of scientific research may exert even more influence across a spectrum, from requiring rapid public uploading of sequences, to requiring delay, to requiring non-disclosure in commercial or sovereign/security cases.[99]

A similar range of beliefs and interests has shaped struggles over cross-sector meta-regulatory norms about data standards and access,[100] and their applications in relation to

96. For a review of literature identifying different ethical issues see generally STEPHANIE JOHNSON AND MICHAEL PARKER, ETHICAL CHALLENGES IN PATHOGEN SEQUENCING: A SYSTEMATIC SCOPING REVIEW (2020) https://wellcomeopenresearch.org/articles/5-119 [https://perma.cc/CSV7-W8NM]. *See also* Arita et al., *supra* note 4, at D123 (explaining that the INSDC supports the Genomic Standards Consortium, which seeks "to build minimum Information about any Sequence (MIxS, where x denotes 'any'; https://gensc.org/mixs/ [https://perma.cc/XXY5-3XYB]) checklists for the collection of rich contextual metadata about biological samples and experimental technologies for a wide variety of genomic data.").

97. *Nature* has expressed and maintained strong support for INSDC. *See* Editorial, *Promoting Best Practice in Nucleotide Sequence Data Sharing*, 7 SCI, DATA, May 2020 (explaining the policy of *Scientific Data* itself as from May 2020: "Authors are required to deposit new non-human sequencing data to an INSDC repository prior to submission, even if the data are already in another open repository. Sample metadata should be deposited alongside sequence data to one of the INSDC Biosample databases. . . We regard sequence data published at Scientific Data and shared through the INSDC repositories as being available for unrestricted use by all researchers in a manner that aligns with principles of open science [reference omitted]").

98. Referred to in WORLD HEALTH ORGANIZATION, *supra* note 14.

99. *See* Genevieve Pham-Kanter et al., *Codifying Collegiality: Recent Developments in Data Sharing Policy in the Life Sciences*, PLOS ONE, Sept. 2014, at 2-7 https://doi.org/10.1371/journal.pone.0108451 [https://perma.cc/ENE3-6KPJ] (discussing influence of funders).

100. 'Meta-regulation' here connotes a set of cross-sector abstract standards intended to influence more specific standards and practices in specific sectors and institutions. For an overview see generally Peter Grabosky, *Meta-Regulation*, *in* REGULATORY THEORY: FOUNDATIONS AND APPLICATIONS 149 (Peter Drahos ed., 2017).

health research and virus GSD. The set of principles of data production and governance known as FAIR (findability, accessibility, interoperability, and reusability)[101] has been incorporated into several systematic sets of data-conduct principles on sequence data sharing, for example by the Global Alliance for Genomics and Health (GA4GH), launched in 2014.[102] While these are expressed functionally, they are also a reaction to the injustice of deadweight loss suffered by other scientists and science beneficiaries where data with valuable potential for others is not used or requires needless costs because of noncompliance with FAIR principles. Other constituencies, however, regard the FAIR principles as addressing one set of justice concerns (and one range of interests) while undervaluing others. The WHO worked over several years on meta-principles for GSD data sharing in a process in which this division came to the fore. With its variegated global set of constituencies and 193 member states, the WHO struggled to bring to finalization a document which in its 2018 version was to be a "Code of Conduct,"[103] but was eventually promulgated in 2022 as the WHO Guiding Principles on pathogen genome data sharing.[104] These include the comment that the FAIR principles on their own are insufficient without incorporating a com-

---

101. *See FAIR Principles*, GOFAIR, https://www.go-fair.org/fair-principles/ [https://perma.cc/4JHK-L82B] (last visited Apr. 7, 2023) (providing further specification of the FAIR Guidelines).

102. *See Framework for Responsible Sharing of Genomic and Health-Related Data*, GLOB. ALL. GENOMICS & HEALTH, https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/ [https://perma.cc/EC42-AYR8] (Dec. 9, 2014) (outlining GA4GH framework which incorporate FAIR principles). GA4GH has also prepared toolkits and normative texts on data sharing. *See* Arita et al., *supra* note 4, D121, D122 (explaining that the ISNDC claims that its framework provided 'the template' for FAIR, and strongly supports FAIR principles).

103. WORLD HEALTH ORGANIZATION, CODE OF CONDUCT FOR OPEN AND TIMELY SHARING OF PATHOGEN GENETIC SEQUENCE DATA DURING OUTBREAKS OF INFECTIOUS DISEASE, Draft (Nov. 2018) [hereinafter WHO Draft Code]. The competing pulls of being specific enough to be useful but open-textured (or vague) enough to secure adoption make it difficult to produce a decisive normative text.

104. This document was developed through consultations but not a formal inter-state negotiating or plenary approval process. *See* WORLD HEALTH ORGANIZATION, *supra* note 14, at iv (indicating that the principles were developed through consultations with experts).

mitment to equity.[105] A collective of indigenous organizations and scholars has developed the CARE Principles for Indigenous Data Governance, which shift from FAIR to emphasize Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE).[106]

One interaction relevant to all of these infrastructural and meta-regulatory contexts is between public *access* to the shared sequence data, and rights then *to use* that data and information developed from it.[107] The period between when data is generated and when it becomes publicly accessible (e.g. through GenBank) is an interval described as data access-latency.[108] The period from data generation to the public having a right to make use of that data (for commercial purposes, or differently for research purposes) is the rights-latency period.[109] For example, patents (which require significant disclosure of the invention to the public in the patents registry) have low information access latency but high rights-of-others-to-use latency.[110] Thus, the HGP set data-access latency at almost zero;[111] GenBank imposes zero rights-latency, but sequence etiquette perhaps does; GISAID imposes some data-access latency as GSD obtained through GISAID cannot be shared except to other GISAID participants; and some rights latency as potential GSD users are obliged to consult with the originating

105. World Health Organization, *supra* note 14, at 5 n.4 ("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.").

106. Stephanie Russo Carroll et al., *Extending the CARE Principles from Tribal Research Policies to Benefit Sharing in Genomic Research*, 13 Frontiers in Genetics (Nov. 11, 2022) at 03.

107. *See* Jorge L. Contreras, *Optimizing Access Policies for Big Data Repositories: Latency Variables and the Genome Commons, in* Big Data Optimization: Recent Developments and Challenges 201 (Ali Emrouznejad ed., 2016) (providing the analytical approach set out in this paragraph).

108. *Id.* at 201. Access latency is not in practice reducible to a binary all-or-nothing access analysis – as suggested by the discussion in the present Article of infrastructural and legal access and use conditions, more nuanced analysis is required where access is available but not to everyone or not without some restrictive conditions.

109. *Id.*

110. *Id.*

111. *Id.*

laboratory.[112] Other data sharing arrangements such as the WHO Pandemic Influenza Preparedness Framework can be analyzed similarly. At the restrictive end of the spectrum, trade secrets law entails substantial data access latency (as a trade secret loses its legal protection if generally disclosed) and substantial rights-latency (knowingly using someone else's trade secret is unlawful).[113] The extension or narrowing of these legal categories and of relevant platform rules may thus have a substantial impact on the availability of pathogen GSD for data sharing.

III. Scientist Cultural-Professional Norms of Attribution and Credit

Science ethics and etiquette issues have been framed as internal to science communities, particularly those communities centered in OECD countries. Some of these ethics and etiquette issues are intensely bound to socio-political phenomena: power gradients, status, identity, and recognition.[114] These phenomena are manifest in Global South-North contexts with regard to GSD.[115] Assessment of science norms and ethics is further complicated by societal variations and by large

112. These are inferences based on the express policies of GISAID and INSDC, summarized above.

113. Jeanne Fromer, *Machines as the New Oompa-Loompas: Trade Secrecy, the Cloud, Machine Learning, and Automation*, 94 NYU L. Rev. 706 (2019).

114. States of Knowledge: The Co-Production of Science and the Social Order (Sheila Jasanoff ed., 2004) ("showing how scientific knowledge both embeds and is embedded in social identities, institutions, representations and discourses"); Robert K. Merton, The Sociology of Science: Theoretical and Empirical Investigations (1973) ("The sociology of science is sometimes defined as a part of the sociology of knowledge, and yet the multifaceted problem of the relations between knowledge and reality (not to speak of the reality of knowledge) is a more general one . . . How do existential, everyday experiences mold the ways in which people conceptualize the world?"). This Article mentions a wide range of science cultures, norms, motivations, and practices, which are highly variable in their purchase, and frequently point in conflicting directions. These are simply a few illustrations of ways in which science norms and cultures may be relevant to understanding GSD and its infrastructures and practices – they are not empirical, nor are they here ordered or reconciled in any way.

115. ThankGod Echezona Ebenezer et al., *Africa: sequence 100,000 species to safeguard biodiversity*, 603 Nature 388–9 (2022) (detailing the effort to launch an African BioGenome Project, involving '109 African scientists (87 of whom work in Africa) and 22 African Organizations' as of March 2022).

variance in practices among different scientific fields, posing challenges in cross-field collaborations.

Professional-cultural norms, as well as perceptions of self-interest, and collective interest among scientists and research units dealing with virus GSD, play a significant role in driving and conditioning data sharing. How best is that role characterized? Lyle Fearnley has argued that the uploading and uses of GSD in these interactions is governed by scientific etiquette, specifically by the set of rules of conduct imposed by GISAID and built into the GISAID EpiFlu databank infrastructure.[116] He gives the example of the A-H7N9 Avian Influenza outbreak among human populations in China in 2013.[117] Virus genome sequences were quickly uploaded by Chinese researchers to GISAID, and their major paper was pending publication when they discovered that a scientist in Japan might be about to publish an analysis of the sequences first.[118] They requested assistance from the GISAID administrators, and a solution was brokered among all parties whereby the Japan paper was delayed by a day so as to appear just after the China paper.[119] The P.R.C., like some other countries, has been reluctant to put biological samples of influenza virus into international circulation,[120] but has regularly supplied genomic sequences. The same pattern seems to have applied for most of the SARS-CoV-2 outbreak. In Fearnley's argument, whereas "viral sovereignty remains rooted in the hierarchical and state-based patterns of international health, sequence etiquette invokes a transnational form, one patterned on the open structure of the internet and other distributed networks" in which scientists in Asia have been able to strategize to "seek fair treatment and the credit that is due to data producers through procedural protocols inscribed in database user agreements."[121]

---

116. Fearnley, *supra* note 1, at 499.

117. *Id.*

118. *Id.* at 499–500.

119. *Id.*

120. *See id.* at 500 (reflecting a PRC lab having been designated a WHO Collaborating Center for Influenza in 2010, and thus under WHO rules an ultimate destination for samples rather than a peripheral supplier of them to a center elsewhere).

121. *Id.* at 501. *See also* Aihwa Ong, *Introduction: An Analytics of Ethics and Biotechnology at Multiple Scales, in* ASIAN BIOTECH: ETHICS AND COMMUNITIES OF FATE 1 (Aihwa Ong & Nancy N. Chen eds., 2010) (discussing Asian countries' biotech policies).

Under this analysis, the platform (in the case of GISAID), or the funders (in the case of the NIH[122]), or other institutions with credibility and direct power, provide and superintend the equilibrium solution to the collective action or trust problem of speedy sequence sharing combined with "first dibs" publication rights and related recognition.[123] Comparable issues about allocating authorship, priority, and status also exist within single labs, within large research organizations, or among different research groups in a hierarchical national science system.[124] These may be covered by formal rules of institutions, overseen and formally or informally operated by specific offices.[125] It is likely that in many of these situations, the acceptance of these rules is buttressed by a shared sense that the rules reflect basic principles of justice or fairness to researchers; or non-rule informal practices may develop and be equilibrated socially.

How is "etiquette" (including sequence etiquette) produced and articulated? Some scientists try to uphold a collective autonomy among scientists, manifested for example in invocations of a 17th century Baconian "invisible college" of scientists who share common standards of judgment, or in more modern nostalgia for a "golden age" of Mertonian ethos & ethics.[126] Mid-twentieth century efforts to observe and theorize these phenomena were influenced by Robert Merton's

---

122. NAT'L INSTIT. OF HEALTH, FINAL NIH POLICY FOR DATA MANAGEMENT AND SHARING, NOTICE NUMBER NOT-OD-21-013 (Jan. 25, 2023), https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html [https://perma.cc/Y82F-8RGL].

123. *Cf.* the theoretical analysis of hierarchical and distributed organizational structures in Matthew R. Zefferman, *Constraints on Cooperation Shape Hierarchical versus Distributed Structure in Human Groups*, 13 SCI. REP. 160 (2023).

124. *See China Delayed Releasing Coronavirus Info, Frustrating WHO*, ASSOCIATED PRESS (June 3, 2020), https://apnews.com/article/united-nations-health-ap-top-news-virus-outbreak-public-health-3c061794970661042b18d5aeaaed9fae [https://perma.cc/VC2E-P79K] (suggesting that such issues within the PRC affected decisions about when and by whom sequences of SARS-CoV-2 would be circulated externally, in the first ten days of January 2020).

125. *See* PAULA STEPHAN, HOW ECONOMICS SHAPES SCIENCE (2012) (detailing how institutions transmit rules of collective action).

126. *See* Mohammad Hosseini et al., *Messing with Merton: The Intersection between Open Science Practices and Mertonian Values*, 30 ACCOUNTABILITY IN RSCH. (2022) [online, no page numbers yet], https://www.tandfonline.

writing on sociology of scientists,[127] focusing on the interaction of earnest motivations to supply scientific research results and insights to a global commons free for all to use, with a reward structure of reputation and prestige which depended on structures for authorial attribution, professional credit, and recognition of priority.[128] A variant drew on Marcel Mauss' social anthropology of the gift-exchange, taken up in Hagstrom's proposition about the enduring significance of the information-for-recognition (or research results-for-esteem) exchange, that "the gift exchange (or the norm of service), as opposed to barter or contractual exchange, is particularly well suited to social systems in which great reliance is placed on the ability of well-socialized persons to operate independently of formal controls."[129] This exchange worked smoothly in the longstanding practices whereby research findings were circulated by publication (or pre-prints), provided there was not leakage or "scooping" during the peer review and editorial process.[130] Conference presentations and similar professional formats also circulated a great deal of knowledge but with some participants more guarded.

Public funding agencies (such as the U.S. NIH) or private philanthropic foundations funding science research, which make grants or issue contracts to research groups and institutions, have to steer in this space. Frequently they prescribe requirements that the results of the work they fund be made available in open access publications and in data reposito-

---

com/doi/full/10.1080/08989621.2022.2141625 (pointing to how some scientists share a nostalgia for Mertonian ethics).

127. Robert K. Merton, *The Normative Structure of Science*, *in* ROBERT K. MERTON, THE SOCIOLOGY OF SCIENCE: THEORETICAL AND EMPIRICAL INVESTIGATIONS 267 (1979).

128. SCIENTIFIC AUTHORSHIP: CREDIT AND INTELLECTUAL PROPERTY IN SCIENCE (Mario Biagioli & Peter Galison eds., 2003).

129. Warren Hagstrom, *Gift Giving as an Organizing Principle in Science*, *in* SCIENCE IN CONTEXT: READINGS IN THE SOCIOLOGY OF SCIENCE 21, 30 (Barry Barnes & David Edge eds., 1982). Hagstrom also commented that "the vast sums spent on space programs, particle accelerators, radiotelescopes, and so forth often seem like a potlatch by the community of nations." *Id.* at 21.

130. *See* Katherine McCain, *Communication, Competition, and Secrecy: the Production and Dissemination of Research-related Information in Genetics*, 16 SCI., TECH. & HUM. VALUES 491 (1991) (building on Hagstrom).

ries.[131] The timing and conditions set in these requirements have been the subject of much deliberation, alteration, and variation. In some funding agencies, key staff may themselves have strong research science backgrounds, and most agencies interact closely with leading research scientists (especially within the same polity) in particular projects and evaluations, so agency prescriptions are often informed by their views as well as other considerations.[132]

The range of interests and the political and existential issues in human genetics greatly exceed those posed by virus genomic research. Nonetheless, the credit, career, and funding issues for researchers are similar enough that efforts to work out science conduct norms (including for data sharing) in human genetics have become a reference point in the etiquette-regulation continuum for model organism, pathogen, and other non-human genomic sequence data.[133] The Human

131. Max Kozlov, *NIH Issues a Seismic Mandate: Share Data Publicly*, NATURE (Feb. 16, 2022), https://www.nature.com/articles/d41586-022-00402-1 [https://perma.cc/J84N-BA75]. And in reverse, as Sabina Leonelli points out, a "major outcome of [large collective] sequencing projects was their success in bringing the scientific importance of activities of data production, dissemination, and integration to the attention of biologists, funding agencies, and governments." SABINA LEONELLI, DATA-CENTRIC BIOLOGY: A PHILOSOPHICAL STUDY 17 (Univ. of Chicago Press, 2016). Leonelli notes as further examples: "The biologists who pioneered the systematic use of model organisms in biology, including T H Morgan in the 1920s (fruit fly), Sydney Brenner in the 1960s (nematode), Maarten Koornneef and Chris Somerville in the 1970s (thale cress), and George Streisinger in the 1980s (zebrafish), shared a similar vision of how research should be conducted. They relentlessly promoted the sharing of ideas, data, and samples as a norm for scientific interactions, including at the prepublication stage—a remarkable feat, particularly in the context of the notoriously competitive culture characterizing biomedical research. This was not simply the result of individual preferences and strong charisma: the wide and free dissemination of data was essential to the scientific success of their research programs, whose ultimate goal was to research organisms as complex wholes." *Id.* at 19.

132. *See* Robert Cook-Deegan et al., *Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance*, 18 ANN. REV. OF GENOMICS AND HUM. GENETICS 389 (2017) (demonstrating the importance of initial communities in many laboratories).

133. *See* JEROME REICHMAN, PAUL UHLIR & TOM DEDEURWAERDERE, GOVERNING DIGITALLY INTEGRATED GENETIC RESOURCES, DATA, AND LITERATURE: GLOBAL INTELLECTUAL PROPERTY STRATEGIES FOR A REDESIGNED MICROBIAL RESEARCH COMMON 406–72 (Cambridge Univ. Press 2016) (detailing early data release policies).

Genome Project (HGP) participation rules required that GSD be released to a shared databank within 24 hours.[134] This was partly to head off risks of patenting, as well as to speed the HGP and avoid needless duplication of the same sequencing work among different groups.[135] As the core funders and coordinators of this massive distributed (multi-lab) operation, the U.S. NIH and its co-funders were able to secure agreement to this, memorialized in 1996 and 1997 in the Bermuda Principles.[136] Once the HGP was nearly complete, a group convened by the Wellcome Trust meeting in 2003 in Fort Lauderdale promulgated a distinction, so that Bermuda-type obligations of rapid pre-publication data release would apply only within those projects whose objective was the creation of datasets or other material "for the broad scientific community" (community resource projects), as contrasted with projects by researchers engaged primarily in hypothesis driven research.[137] Pushback against any subsequent generalization of the HGP GSD sharing rules was strong, culminating in the 2014 NIH

---

134. *Summary of Principles Agreed Upon at the First International Strategy Meeting on the Human Genome Sequencing*, HUMAN GENOME PROJECT INFORMATION ARCHIVE 1990–2003, (February 25–28, 1996), https://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml [https://perma.cc/GUR5-8W66]. The HGP is an influential referent because it deployed large financial resources and brought together major science funders and a very large number of genetic sequence scientists and laboratories, as well as infrastructures and publishers; through these different vectors its norms and practices thereafter diffused widely into non-human genomic data work.

135. Jorge Contreras, *Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons,* 12 MINN. J. L., SCI, & TECH. 61, 82 (2011).

136. *Summary of Principles*, *supra* note 134. For discussion see Contreras, *supra* note 135, at 64. The 1997 meeting set forth principles concerning sequence claims and etiquette, as well as standards for sequence quality, and rules on sequence submission and annotation. *See Summary of the Report on the Second International Strategy Meeting on Human Genome Sequencing*, HUMAN GENOME PROJECT INFORMATION ARCHIVE 1990–2003 (February 27th–March 2nd, 1997), https://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml [https://perma.cc/WBA3-UEXD].

137. THE WELLCOME TRUST, SHARING DATA FROM LARGE-SCALE BIOLOGICAL RESEARCH PROJECTS: A SYSTEM OF TRIPARTITE RESPONSIBILITY (2003) https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf [https://perma.cc/Q2XH-ZKWM]. *See* Contreras & Knoppers, *supra* note 90*,* at 437 (noting that many scientific research areas include community resource projects, often dependent on coordination and data sharing organized by a central funder).

policy that allows the sequencing lab and/or the sample-originating lab a 6–12 month window to first prepare their publications in relation to human GSD, and freedom to hold off making non-human GSG generally available until publication occurs.[138] The tip back from a norm of rapid pre-publication sequence sharing, to a privileging of more traditional methods of establishing priority and securing recognition through publication authorship, was itself vigorously contested.[139]

Contemporary genomic sequencing of a pandemic virus is a networked enterprise, but it is not orchestrated through centrally-directed "big science" with thousands of scientists in numerous countries all performing an allocated task and a lead group of planners (often also funders) at the core. It thus differs from the Human Genome Project,[140] and from what was attempted by "international unions" of the late nineteenth and early twentieth century in large closely-coordinated projects.[141] Nonetheless, the operation of collection, sequenc-

---

138. Cook-Deegan et al., *supra* note 132, at 399. This article and Contreras & Knoppers, *supra* note 90, each provide useful overviews of this whole area, focused particularly on the US. *See also* REICHMAN, UHLIR & DEDEURWAERDERE, *supra* note 133, at 406–413 (discussing 'Early Release Policies to Manage the Deluge of Genomic Reference Data'). Leading journals publishing scientific papers now typically require that the underlying data be made available to others, often in an open repository, to facilitate testing of published results and replication or critique. Efforts have been made over many decades to marshal the editors of the leading journals into a role as collective regulators requiring open availability of data, with mixed success. For example, enlisting editors was part of initial proposals for GenBank in the early 1980s. Strasser, *supra* note 29.

139. *See* Rudolf I. Amann et al., *Toward Unrestricted Use of Public Genomic Data: Publication Interests Should Not Limit Access to Public Data*, 363 SCI. 350 (2019) (discussing the disagreements regarding the public sharing of data).

140. BIG SCIENCE: THE GROWTH OF LARGE-SCALE RESEARCH (Peter Galison & Bruce Hevly eds., 1992); ROBERT COOK-DEEGAN, THE GENE WARS: SCIENCE, POLITICS, AND THE HUMAN GENOME (1994); EUGENE THACKER, THE GLOBAL GENOME: BIOLOGY, POLITICS, AND CULTURE (2005).

141. *See, e.g.*, Peter Galison & Lorraine Daston, *Scientific Coordination as Ethos and Epistemology*, *in* INSTRUMENTS IN ART AND SCIENCE: ON THE ARCHITECTONICS OF CULTURAL BOUNDARIES 296, 308–314 (Helmar Schramm, Ludger Schwarte & Jan Lazardzig eds., 2008) (describing effort to determine the shape of the earth through a coordinated global mapping of local gravity and trigonometry by the Europäischen Gradmessung (1862–c. 1914), forerunner of the non-governmental International Union of Geodesy and Geophysics (1919–present, based in Potsdam)).

ing, uploading, and wide use of the sequences by other scientists is one of large-scale transnational science collaboration.

The WHO has been centrally involved in managing these dilemmas involving pathogen virus GSD. In the mid-2000s, the WHO sought to broker a compromise for influenza virus sequences: researchers would be incentivized to continue to share GSD relating to H5N1 viruses with pandemic potential via the Los Alamos Influenza Sequence Databases (ISD),[142] as access to the sequences would be restricted to a WHO-designated group of laboratories. These laboratories would be obliged to consult with the original research laboratory prior to any publication, thus reducing the risk of the researchers being "scooped" or someone else claiming intellectual property rights.[143] Disagreement with this limited access arrangement led Ilaria Capua and other scientists to establish what became GISAID.[144]

The current etiquette, norms, and arrangements have secured significant sharing of virus GSD, but many scientists and laboratories do not upload all (or in some cases, any) virus GSD to accessible databanks, at least until they are required to in support of publication (or patent applications) or regulatory objectives. Other explanatory factors include concerns about intellectual property claims they could make, or which others may use the GSD to make; national and international interests in ensuring a flow of benefits in return for providing GSD access; and national-scale sequencing coordination as well as geopolitical positioning and agendas.[145] These issues are addressed in the next section.

---

142. Los Alamos National Laboratory, *Los Alamos Introduces New Influenza Database,* SCI. DAILY (Jul. 22, 1998).

143. Fearnley, *supra* note 1, at 492 (explaining that only the fifteen or so laboratories in the WHO Global Influenza Surveillance and Response System (GISRS) would have the access password, and all would be bound to consult the originating laboratory before any publication).

144. *See id.* at 493–95 (discussing the formation of GISAID).

145. All of these factors were mentioned in interviews with French virus scientists conducted for a French newspaper probing the disparity suggested by U.K. researchers having uploaded 35,965 genomic sequences to GISAID in a period from March–August 2020 while French researchers had uploaded 559 (including three cat virus sequences). David Larousserie, *Covid-19: les chercheurs français peu partageurs des séquences génétiques*, LE MONDE (Aug. 31, 2020, 6:30 AM) https://www.lemonde.fr/sciences/article/2020/

IV. State National/International Law of Virus Sovereignty, Biological-Materials Benefit Sharing, and Data Sharing

A different kind of attributive justice argument has been made by governments of developing states—with variants articulated by indigenous peoples and other communities—that these states or indigenous peoples and local communities have sovereign rights in relation to biological samples and other genetic resources originating in their territories or in their communities. Where they have direct control over the sample or genetic resource, they may endeavor to set conditions on access to it, including respect for cultural protocols, acknowledgement of use in downstream outputs utilizing the resource, and economic benefits; hence the standard terminology of access-benefit exchange.[146] Similar claims are also made by states and indigenous groups where the original resource has already been taken from their territory—and is held in a repository in another country—with more leverage if the extraction occurred after the entry into force of the UN Convention on Biodiversity (CBD) of 1992.[147] The CBD is by no means the only embodiment of a broad access-benefit exchange principle in international treaty law. Among the other relevant precedents were the 1982 UN Convention on the Law of the Sea (UNCLOS) with regard to coastal state rights to control marine scientific research as well as coastal state resources rights in the continental shelf and in the exclusive economic zone out to 200 nautical miles from the coast,[148] which was supplemented by the 1995 treaty on straddling fish stocks.[149]

08/31/covid-19-les-chercheurs-francais-peu-partageurs-des-sequences-genetiques_6050428_1650684.html [https://perma.cc/KAB2-XQRH].

146. A database of some actual and model agreements curated by the World Intellectual Property Organisation (WIPO) is *Biodiversity-related Access and Benefit-sharing Agreements,* WIPO https://www.wipo.int/tk/en/databases/contracts/ [https://perma.cc/QZA6-PZSZ] (last visited May 10, 2023).

147. Convention on Biological Diversity (Rio de Janeiro, June 5, 1992) 1760 U.N.T.S. 79, *entered into force* 29 Dec. 1993.

148. United Nations Convention on the Law of the Sea (Dec. 10, 1982) 1833 U.N.T.S. 3, 21 I.L.M. 1261 (1982), *entered into force* Nov. 16, 1994 [hereinafter UNCLOS].

149. Agreement for the Implementation of the Provisions of the United Nations Convention on the Law of the Sea of 10 Dec. 1982 relating to the Conservation and Management of Straddling Fish Stocks and Highly Migra-

The CBD, to which all UN member states are parties with the exception of the United States, has been the most influential platform for the articulation and further development of an access-benefit approach to (non-human) genetic resources. The 2010 Nagoya Protocol (NP) to the CBD,[150] to which 140 or more states are parties,[151] provides in much more specific terms the treaty rules and rule-making structures, and a basis for detailed national (implementing) legislation and administrative action, to operationalize this principle.[152] Concerns about intellectual property rights claims by external exploiters of resources, and some concerns centered on non-pecuniary issues of fair credit and attribution to local scientists and agencies for their initial work in collecting samples and isolating a virus, have been factors influencing developing country governments in their arguments for national legislation and international agreement on access-benefit regimes. These restrict both access of foreign researchers and sharing of data abroad, unless a suitable benefit sharing arrangement is in place.[153]

The WHO's Pandemic Influenza Preparedness (PIP) Framework, adopted in 2011 after tortuous negotiations, simi-

---

tory Fish Stocks (New York, Aug. 4, 1995) UN Doc. A/CONF.164/37, *entered into force* 11 Dec. 2001. *See also* Margaret Young & Andrew Friedman, *Biodiversity Beyond National Jurisdiction: Regimes and Their Interaction*, 112 AM. J. INT'L L. UNBOUND 123 (2018) (discussing international instruments for oceans governance).

150. For an early overview see ELISA MORGERA, ELSA TSIOUMANI & MATTHIAS BUCK, UNRAVELING THE NAGOYA PROTOCOL: A COMMENTARY OF THE PROTOCOL ON ACCESS AND BENEFIT-SHARING TO THE CONVENTION ON BIOLOGICAL DIVERSITY (2014).

151. As of July 30, 2023, 140 states have ratified the Nagoya Protocol. A further 15 states signed but have not (as yet) ratified. *See Parties to the Nagoya Protocol*, CONVENTION ON BIOLOGICAL DIVERSITY, https://www.cbd.int/abs/nagoya-protocol/signatories/ [https://perma.cc/2CHD-GRMM] (last visited July 1, 2023).

152. For more information, see the *Access and Benefit-sharing Clearing-House*, CONVENTION ON BIOLOGICAL DIVERSITY, https://absch.cbd.int/en/ [https://perma.cc/97YJ-BU2T] (last visited May 10, 2023).

153. *See* Margo Bagley, *"Just" Sharing: The Virtues of Digital Sequence Information Benefit-Sharing for the Common Good*, 63 HARVARD INT'L L.J. 1 (2022) (discussing digital sequence information benefit-sharing); *see also* Michelle Rourke, *Viruses For Sale: All Viruses are Subject to Access And Benefit-Sharing Obligations under the Convention on Biological Diversity*, 39 EUR. INTELL. PROP. REV. 79 (2017) (discussing benefit sharing obligations under the CBD).

larly seeks to incentivize biological sample sharing.[154] The negotiations were precipitated by Indonesia's withdrawal from H5N1 virus sample sharing when in 2006 a biological sample containing an avian influenza virus, sent by Indonesia to a WHO laboratory abroad and supplied from there to an Australian company, was used in a patent on the path to developing a vaccine.[155] The PIP framework develops Materials Transfer Agreements between the originating country (of the sample) and the WHO lab receiving the sample, after which a further MTA governs any transfer from the WHO lab system to any third party.[156] The MTAs are supposed to ensure that a portion of relevant vaccines, and of pecuniary benefits, flow back to the originating country if the sample results in a successful vaccine or biotech product, with a wider needs-based element also incorporated.[157] However, there are doubts that the MTAs actually achieve that in legal terms,[158] and so far the PIP Framework seems in practice to have been more significant as a model than as the basis for large benefit flows. It has generally been assumed that the PIP Framework when adopted was limited to physical samples and did not directly apply to GSD,[159] but as GSD has become a more central issue developing countries, alignment will increasingly be sought between the PIP Framework and GSD regimes and infrastructures.

The most important demarche was Indonesia's in 2006–2007.[160] As a major normative aim was to reject the argument that someone outside Indonesia might have legal rights to a virus extracted from Indonesia—in this case patent rights to the virus or the vaccine—Indonesia and other countries ar-

---

154. For more on the PIP Framework, *see* Michelle Rourke, *Access by Design, Benefits if Convenient: A Closer Look at the Pandemic Influenza Preparedness Framework's Standard Material Transfer Agreements*, 97 MILBANK Q. 91 (2019).

155. Shahar Hameiri, *Avian Influenza, "Viral Sovereignty", and the Politics of Health Security in Indonesia,* 27 PAC. REV. 333, 334–7 (2014).

156. Rourke, *Access by Design, supra* note 154, at 94.

157. *Id.* at 95.

158. *Id.* at 97-108.

159. Lawrence O. Gostin et al., *Virus Sharing, Genetic Sequencing, and Global Health Security,* 345 SCI. 1295, 1295–96 (2014).

160. For a contemporaneous account outlining some of the arguments as they were understood among Indonesian scientists and policymakers, see Endang R. Sedyaningsih et al., *Towards Mutual Trust, Transparency and Equity in Virus Sharing Mechanism: The Avian Influenza Case of Indonesia*, 37 ANNALS ACAD. MED. SING. 482 (2008).

ticulated a claim to "virus sovereignty." In international terms, this was an argument that imperium defeats an externally grounded dominium. It reflected deep concern in many developing countries about the exclusions and inequitable distributions wrought by international intellectual property rules,[161] combined with episodes of uncompensated and unlicensed exploitation of their resources by companies from richer countries which they characterize as "biopiracy." The working out of intellectual property law in relation to genomics, especially patentability, and sequences has been complex.[162] It is true that IPR right-holders may use their rights in a pro-data sharing fashion, and that their plans may in some cases align with interests of developing countries. Some sequences were made public to prevent other sequencers from patenting them, particularly when the law allowed patenting of genomic sequences with unknown functions. In other cases patents were sought on sequenced genetic materials, processes, or treatments associated with them, as a defensive strategy to pre-empt commercial or external entities securing such patents.[163] This is what Erasmus University in the Netherlands claims to have done with regard to the Middle East Respiratory Syndrome (MERS) virus in 2012, which it had isolated in a sample sent from Saudi Arabia.[164] Nonetheless, the patent caused considerable dis-

161. *See generally* ROCHELLE DREYFUSS AND GRAEME DINWOODIE, A NEOFEDERALIST VISION OF TRIPS: THE RESILIENCE OF THE INTERNATIONAL INTELLECTUAL PROPERTY REGIME (Oxford Univ. Press 2012) (providing a detailed and critical overview of TRIPS). *See also* Helen Gubby, *Is the Patent System a Barrier to Inclusive Prosperity? The Biomedical Perspective*, 11 GLOB. POL'Y 46 (2020) (arguing that 'the whole biomedical sector should be taken out of the ambit of the patent system' and innovators (including Pharma) funded by state grants, as well as rewards and prizes); PATENTS ON LIFE: THROUGH THE LENSES OF LAW, RELIGION, AND GLOBAL JUSTICE (Thomas C. Berg, Roman Cholij & Simon Ravenscroft eds., 2019); PATENT LAW IN GLOBAL PERSPECTIVE (Ruth Okediji & Margo Bagley eds., 2014); JORGE CONTRERAS, THE GENOME DEFENSE (2022).

162. For an overview see Contreras & Knoppers, *supra* note 90, at 443–446.

163. *See generally* Jorge Contreras, *Public Licenses: Open Source, Creative Commons and IP Pledges, in* INTELLECTUAL PROPERTY LICENSING AND TRANSACTIONS: THEORY AND PRACTICE 592 (Jorge Contreras ed., 2022) (explaining that such patents could then be made freely available for all to use under a public license).

164. CENTER FOR GLOBAL HEALTH SCIENCE AND SECURITY, GEORGETOWN UNIVERSITY, MERS-CoV DATA SHARING CASE STUDY REPORT (Nov. 2018),

quiet to the relevant Saudi elite, and likely slowed information sharing regarding MERS.[165]

Sovereignty claims are also interwoven with responsibility for risks from bio-samples, whether safety risks from accidental exposure and storage degradation, to risks from deliberate release or weaponization. The problem of storing and disposing of samples is a large one in post-epidemic management, manifest after Ebola epidemics in West Africa from 2014–2016 and then in the D.R.C. and neighboring countries from 2018–2019. A WHO review in 2016 estimated there were about 162,000 EBOV samples it could identify, of which ten percent were positive samples. 42,000 of these samples were stored in laboratories outside the affected countries.[166] Of the remaining 120,000 identified samples, approximately 28,000 were stored in Liberia, approximately 58,000 in Guinea, and approximately 33,000 in Sierra Leone.[167]

The "viral sovereignty" idea waned somewhat under criticism from scientists and from wealthy OECD countries, and perhaps also due to recognition of the irony of asserting sovereignty over a non-living pathogen which might become a mass killer worldwide.[168] Genomic sovereignty has become a more popular or palatable alternative among advocates and government leaders in developing countries,[169] with support also in China.[170] Genomic sovereignty may be linked also to "a postgenomic turn that brings with it the reanimation of group dis-

---

https://www.glopid-r.org/wp-content/uploads/2019/07/Georgetown_MERS-Case-Study-Report-1.pdf [https://perma.cc/9SF8-EQVY].

165. *Id.*

166. Joshua Teperowski Monrad & Rebecca Katz, *Biosecurity, Biosafety, and the Management of Dangerous Pathogens for Public Health*, *in* VIRAL SOVEREIGNTY AND TECHNOLOGY TRANSFER: THE CHANGING GLOBAL SYSTEM FOR SHARING PATHOGENS FOR PUBLIC HEALTH RESEARCH 100 (Sam Halabi & Rebecca Katz eds., 2020).

167. *Id.* (noting that the WHO launched a voluntary Joint External Evaluations Tool (JEE) in 2016, used in over 100 countries, involving site visits and external evaluations of core capacities for health security).

168. A useful review of its arc is Michelle Rourke, *Restricting Access to Pathogen Samples and Epidemiological Data: A Not-So-Brief History of "Viral Sovereignty" and the Mark it Left on the World, in* INFECTIOUS DISEASES IN THE NEW MILLENNIUM 167 (Mark Eccleston-Turner & Iain Brassington eds., 2020).

169. *Id.*

170. LYLE FEARNLEY, VIRULENT ZONES: ANIMAL DISEASE AND GLOBAL HEALTH AT CHINA'S PANDEMIC EPICENTER (Duke Univ. Press 2020).

tinctions along local, national, and continental lines."[171] The conversion of genomic resources into highly mobile GSD poses large challenges to a genomic sovereignty program.[172] Any governance arrangements and legal entitlements relating to GSD necessarily interact with, and are likely to depend on, GSD infrastructures as a means of direct regulation as well as a means of supporting state or international legal regulation. Multilateral processes and particular states have increasingly engaged in direct or indirect dialogues with GSD infrastructural governance entities, pressing them to do more to facilitate capture of benefits for sharing under both attributive entitlements and proposed distributive multilateral arrangements. One issue is the practice of sequences being uploaded and circulated without tags indicating the country from whence the sample originated or the date of collection, which the major infrastructures and journals largely tolerated. However, the value of this information for sovereignty and ABS claims overlaps with the value of extensive metadata in scientific uses of the GSD, and for reasons of science and socio-political pressure both GISAID and INSDC have moved toward requiring it.[173] In the case of INSDC, which had faced considerable

171. Amy Hinterberger & Natalie Porter, *Genomic and Viral Sovereignty: Tethering the Materials of Global Biomedicine*, 27 PUB. CULTURE 361 (2015) (footnotes omitted).

172. This dichotomy between the law of physical materials and the law of information is drawn ably, but perhaps more sharply than practice supports, in Lawson et al., *supra* note 40, and in Charles Lawson & Michelle Rourke, *Digital Sequence Information as a Marine Genetic Resource Under the Proposed UNCLOS Legally Binding Instrument*, 122 MARINE POL'Y 1 (2020).

173. For example, INSDC Spatio-temporal annotation policy announced on 18 November 2021 states (insdc.org/policy) [https://perma.cc/D538-UDUD]: "INSDC aims to increase significantly the number of sequences for which the origin of the sample can be precisely located in time and space. We will achieve this through harmonisation of accurate geographical annotation and time of collection information. Our ultimate goal is to ensure spatio-temporal annotation is collected for all new incoming sequences by the end of 2022. Over the next year, you can expect to see INSDC databases starting to put in place additional requirements for new sequence submissions. For example, in future we expect to require that all new submissions include for each sample: The country or region (from https://www.insdc.org/country.html) where the sample was collected, using standardised country names from a controlled list[; and] The collection date of the sample, recording at least the year of collection." INSDC, SPATIO-TEMPORAL ANNOTATION POLICY (Nov. 18, 2021), https://www.insdc.org/news/

pressure from developing countries to do more toward associating sequences with countries of sample origin for future ABS purposes, the explicit welcome of INSDC's concession in a formal 2022 decision of the CBD Conference of the Parties was indicative of the INSDC and the CBD majority, each moving in response to one another.[174]

Indonesia's situation was unusual in that there was a substantial prospect that the next avian flu pandemic virus would emerge among avian and human populations in Indonesia rather than elsewhere.[175] There was thus a location-specific dimension which the Indonesian government could properly represent, and to some extent the PIP Framework proceeds on the premise that location-specific characteristics often exist in relation to zoonotic influenza viruses which may induce human pandemics. However, while some zoonotic viruses are localized in this way, many genetic resources are found in multiple countries, creating arbitrage opportunities for companies and even incentives to disguise a specific national origin in order to evade strong regulations or reduce payments.[176]

A different approach to access-benefit sharing (ABS) is exemplified by efforts to organize ABS in areas "beyond national jurisdiction", including the deep seabed provisions of UNCLOS,[177] and the 2023 UNCLOS-related instrument on biodiversity in marine areas beyond the existing limits of national jurisdiction.[178] Many of the relevant resources are not

---

spatio-temporal-annotation-policy-18-11-2021/ [https://perma.cc/7YG6-JCR9].

174. Decision Adopted by the Conference of the Parties to the Convention on Biological Diversity, U.N. Doc. CBD/COP/DEC/15/9 preamble (Dec. 19, 2022).

175. Kathryn E. Lafond et al., *Seasonal Influenza and Avian Influenza A(H5N1) Virus Surveillance among Inpatients and Outpatients, East Jakarta, Indonesia, 2011–2014*, 25:11 Emerging Infectious Diseases 1 (2019).

176. Rourke, *supra* note 41.

177. UNCLOS, *supra* note 148, arts 136–141. Elise Morgera, *Fair And Equitable Benefit-Sharing: History, Normative Content and Status In International Law* (BENELEX Working Paper No. 12, rev. June 2018).

178. Intergovernmental Conference on an International Legally Binding Instrument Under the United Nations Convention on the Law of the Sea, *Agreement Under the United Nations Convention on the Law of the Sea on the conservation and sustainable use of marine biological diversity of areas beyond national jurisdiction*, U.N. Doc. A/CONF.232/2023/4 (June 19, 2023) (not yet in force) [hereinafter *Agreement under UNCLOS on marine biological diversity of*

yet known, or are known but their potential economic value and uses not known. This aspect opens the possibility of negotiating treaties in which all participants become better off and none worse off (Pareto-improvements), at least vis-à-vis current activity. Such equilibria are much easier to negotiate than changes in current distributions of rights and uses, which require compensation and activity changes (or grandparenting-in existing uses) in order to achieve an improved Kaldor-Hicks equilibrium.[179]

Article 10 of the Nagoya Protocol explicitly mentions the potential need for a multilateral benefit sharing fund in the context of transboundary genetic resources and associated traditional knowledge, or for situations in which it is not possible to obtain prior informed consent.[180] In such cases meeting ABS and Prior Informed Consent obligations through payments into a multilateral benefit sharing fund, might be more efficient, more just, and relatively more likely to elicit compliance from commercial entities otherwise tempted to evade all payment.[181] More broadly, it is increasingly argued that creating a substantial Multilateral Fund might avoid both the risk of originating states limiting access and the risk that researchers apprehensive about liability might refrain altogether from using genomic resources to which the Nagoya Protocol applies. The International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), administered by the United Nations Food and Agriculture Organization (FAO), provided a precedent for such a multilateral pooled approach.[182] The 2023 UN agreement on marine biodiversity beyond national jurisdiction (BBNJ) adopted a comparable multilateral benefit-sharing mechanism.[183]

---

areas beyond national jurisdiction]. For earlier commentary, see Lawson & Rourke, supra note 172.

179. Lionel Robbins, *Economics and Political Economy*, 71 AM. ECON. REV. 1, 6 (1981).

180. ELISA MORGERA, ELSA TSIOUMANI & MATTHIAS BUCK, UNRAVELING THE NAGOYA PROTOCOL: A COMMENTARY OF THE PROTOCOL ON ACCESS AND BENEFIT-SHARING TO THE CONVENTION ON BIOLOGICAL DIVERSITY 197–208 (2014).

181. ACCESS AND BENEFIT SHARING OF GENETIC RESOURCES, INFORMATION AND TRADITIONAL KNOWLEDGE (Charles Lawson, Michelle Rourke & Fran Humphries eds., 2023).

182. MORGERA, *supra* note 180, at 205–6.

183. Agreement under UNCLOS on marine biological diversity of areas beyond national jurisdiction, *supra* note 178.

The difficulties of tying dematerialized GSD to both a country of biological origin and to a traceable commercial product, even if sometimes possible to overcome, were sufficiently severe to give special impetus to proposals for a multilateral fund as a means of benefit-sharing obligations where the key economic driver was GSD. Some third-party commercial uses of genetic resource information from publicly accessible databases can be analogized to a transboundary situation or to another situation where it is not feasible to obtain consent.[184] Such cases may arise where synthetic biology researchers use fragments of DNA sequences from many different species in designing new biosynthesis pathways to generate new or enhanced compounds.[185]

Whereas these incipient multi-country and global-benefit moves open the paths to including distributive justice principles within the allocation mechanism—as had been attempted also in the 1982 UNCLOS deep sea-bed minerals regime—the CBD and NP primarily allocate rights to genetic resources based on attributive principles, using traditional international law allocations of jurisdictional authority. The starting point here relates to rights over resources derived from sovereignty over the relevant territory. These use the treaty to create obligations as between states, and for the most part envisage that domestic legislation in the resource providing country and in the country where the resource is eventually used (or in the country regulating the commercial user) will impose the ABS structure on private sector uses. Taking the CBD-NP as the primary case, several features may be noted.

The CBD and the NP apply to "genetic resources," and the ABS arrangements are most easily read as focused on physical materials (biological samples etc.).[186] A separate strand of

184. Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits arising from their Utilization to the Convention on Biological Diversity, 13th preambular recital, Oct. 29, 2010, 3008 U.N.T.S. 24.

185. *See* CONVENTION ON BIOLOGICAL DIVERSITY, TECHNICAL SERIES ON SYNTHETIC BIOLOGY OF THE SECRETARIAT OF THE CONV. ON BIOLOGICAL DIVERSITY, No. 100 (2022) (describing a case where researchers recreated an extinct infectious horsepox virus from chemically synthesized DNA fragments).

186. As early as 1995 the CBD COP stated that *human* genetic resources are outside the purview of the CBD (notwithstanding that the scope of application of CBD does not in itself exclude them). Report of the Second Meet-

CBD-NP focuses on the overall objective of promoting exchanges of information including research results, specialized knowledge, and traditional knowledge. The COP/MOP began systematically to consider what it decided to call "digital sequence information on genetic resources" (hereafter "DSI") from 2016 onward.[187] Comments by states indicated no consensus on whether DSI/GSD are already included within the scope of CBD-NP ABS obligations.[188] However CBD-NP mainly sets a floor rather than a ceiling. Provided it is not contrary to other treaties or international law rules, a state can legislate certain restrictions on external transmission or use of GSD as a condition for access to biological samples in its territory, and potentially assert some extraterritorial reach.[189] But other international obligations may constrain the state from exercising this power. One such treaty is the International Health Regulations of 2005, adopted under the auspices of the WHO.[190] The

---

ing of the Conference of the Parties to the Convention on Biological Diversity, at para. 2, UNEP/CBD/COP/2/19 Decision II/11 (1995). It also seems to be well established that ordinarily the CBD and NP do not generally apply to samples that were already outside the country of origin at the date of entry into force of the CBD in 1993.

187. *See generally* reports and documents at CONVENTION ON BIOLOGICAL DIVERSITY, https://www.cbd.int (last visited Apr. 23, 2023).

188. Elizabeth Karger et al., *Digital Sequence Information on Genetic Resources (DSI): An Introductory Guide for African Policymakers and Stakeholders* 7 (2019) https://unctad.org/system/files/official-document/ditc-ted-05052020-BioTradeSSC-DSI.pdf [https://perma.cc/3NF4-A2PW]; AFRICAN GROUP OF NEGOTIATORS' ON BIODIVERSITY-AD HOC GROUP ON DIGITAL SEQUENCE INFORMATION, DIGITAL SEQUENCE INFORMATION ON GENETIC RESOURCES: SUBMISSION OF VIEWS AND INFORMATION ON TERMINOLOGY, SCOPE, AND DOMESTIC MEASURES ON ACCESS AND BENEFIT SHARING 2 (2019), https://www.cbd.int/abs/DSI-views/2019/AfricanGroup-DSI.pdf [https://perma.cc/Q8XT-HNRH]; Bagley, *supra* note 153.

189. Margo Bagley et al., *Fact-finding Study on How Domestic Measures Address Benefit-sharing Arising from Commercial and Non-commercial Use of Digital Sequence Information on Genetic Resources and Address the Use of Digital Sequence Information on Genetic Resources for Research and Development* 12 (2019) https://www.cbd.int/abs/DSI-peer/Study4_domestic_measures.pdf [https://perma.cc/4BCT-SU83] (discussing various domestic measures states have used to address DSI). Bagley, *supra* note 153, at 13.

190. WHO Assembly Res. WHA58/2005/REC/1, at 7-9 (May 23, 2005) (subsequent amendments did not materially affect the issues discussed here). *See also* Gian Luca Burci & Stefania Negri, *Governing the Global Fight Against Pandemics: The WHO, the International Health Regulations, and the Fragmentation of International Law*, 53 N.Y.U. J. INT'L L. & POL. 501 (2021) (critiquing the International Health Regulations governance model).

relationship between CBD-NP and the International Health Regulations remains contentious. Each is a legally binding text.[191] Of the CBD's 196 states parties (the United States is the only UN member not party to the CBD), 140 are bound by the NP. Nearly all of these are members of the WHO and bound by the IHR. They are thus obliged to two major international law agreements pulling in opposite directions.

A fragmented mosaic of such rules, with a seasoning of legal and political uncertainty, is viewed by many scientists as highly undesirable for academic research, as well as raising objections from biotech and pharma industries. At COP15, which concluded in December 2022, the states parties to the CBD adopted a diplomatically worded decision on DSI which, while preserving a basis for many unreconciled differences in the positions taken by different states, seemed to portend a substantial acceptance of a pooled multilateral approach as a distinctive solution for benefit-sharing issues in relation to DSI.[192] The COP decided to establish a multilateral mechanism for benefit-sharing from the use of DSI on genetic resources, including a global fund, and prescribed a two-year time-bound process to further develop and operationalize the mechanism.[193] The COP Decision encourages the depositing

191. *See generally* WHO Secretariat, Implementation Of The Nagoya Protocol And Pathogen Sharing: Public Health Implications (2017), https://cdn.who.int/media/docs/default-source/documents/nagoya-protocol/nagoya-full-study-english.pdf?sfvrsn=EC2ab49d_12&download=true [https://perma.cc/YLS9-4BAE] (study on the implementation of the Nagoya Protocol in consideration of inter alia the International Health Regulation).

192. Several points of divergence were expressed via clause 11, in which the COP "Agrees that the approach set out in this decision to fair and equitable benefit-sharing from the use of digital sequence information on genetic resources does not affect existing rights and obligations under the Convention and the Nagoya Protocol, including, as applicable, those related to traditional knowledge and the rights of indigenous peoples and local communities, and is without prejudice to national access and benefit-sharing measures." Thus national legislation regulating GSD and setting conditions on its generation and use might well continue. Other contested matters were explicitly left for future resolution, including the relationship of the CBD multilateral mechanism to the Nagoya Protocol. Conference of the Parties to the Convention on Biological Diversity (15th Meeting), Decision 15/9 on Digital Sequence Information on Genetic Resources, U.N. Doc. CBD/COP/DEC/15/9, at 3 (Dec. 19, 2022).

193. *Id.* at 4.

of more DSI in public databases "with appropriate information on geographical origin and other relevant metadata," while also recognizing "that tracking and tracing of all digital sequence information on genetic resources is not practical."[194]

The CBD COP15 Decision does not draw any distinction between pathogen GSD and other DSI/GSD, giving rise to the inference that the global fund mechanism would include benefit payments relating to pathogen sequencing. This does not in itself appear to be contrary to the IHR, provided data is shared publicly and promptly in forms suitable for analysis and not withheld for benefits reasons. The merger of scientist-credit and national access-benefit sharing claims helps to explain the intensity of feeling among scientists and policymakers in locations where they have been subject to extraction of virus genetic material with little attribution and little return benefit. This combination of attributive justice claims, but with a strongly needs-based or distributive justice sensibility, is captured in a report from a microbiologist who worked at a molecular-biology lab in Guinea during the 2014–16 Ebola outbreak, and who commented that most of the key papers on that outbreak were written by European and North American scholars who had ended up with the data: "We handled many samples and assumed they belonged to the country . . . But all the samples were shipped out."[195]

A human rights approach shifts the accent somewhat from strictly attributivist justice to take account of distributive justice dimensions. Alexandra Phelan makes a case that this shift "decouples the directly transactional relationship, for an approach that prioritizes global health and conceptualizes access and benefits as intrinsically linked through a right to health lens."[196] While human rights are a normative beacon in work on health and science and provide tremendous impetus to advocacy and mobilized contestation against abuses, scien-

194. *Id.* at 2.

195. Amy Maxmen, *Ebola Researchers Hunt for Cures in a War Zone*, 572 NATURE 16, 17 (2019). On West Africa, see also Emmanuel Freudenthal, *Ebola's Lost Blood: Row over Samples Flown out of Africa as "Big Pharma" Set to Cash In*, THE TELEGRAPH (Feb. 6, 2019).

196. Alexandra Phelan, *Human Rights Implications of Pathogen Sharing and Technology Transfer, in* VIRAL SOVEREIGNTY AND TECHNOLOGY TRANSFER: THE CHANGING GLOBAL SYSTEM FOR SHARING PATHOGENS FOR PUBLIC HEALTH RESEARCH 120, 128–129 (Sam Halabi & Rebecca Katz eds., 2020).

tists and developing countries may have doubts that the abstract norms and universal institutions of human rights will have impact on distributive justice to meet all their concerns. The proposition that "the enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being . . ."[197] (as the WHO Constitution Preamble puts it) is almost universally accepted but its leverage is limited in specific governance and infrastructural contexts, in which not only technologies and organizations but also social practices and civic epistemologies configure in kaleidoscopic ways and move and transform unexpectedly. The right to share in the benefits of science,[198] which has begun to receive some flickering attention after long neglect,[199] might be thought to have potential significance for virus GSD but has not (yet) been energized in this direction.[200]

Arguments for universal human rights and global equity generate contestation against the expansion of attribute justice claims which may undercut distributive justice claims. These distributive justice claims prioritize flows of vaccines and health care to poor people and to needy societies over claims of originating states or sequence-generators.

---

197. Constitution of the World Health Organization 1946, 1 CHRON WORLD HEALTH ORGANIZATION 29 (1947), Preamble, second recital.

198. The statement on the right of everyone 'to share in scientific advancement and its benefits', included as Article 27 of the Universal Declaration of Human Rights in 1948, was a convergence of language although not substance between the Soviet Union and the United States and its allies. The more detailed Article 15 of the ICESCR 1966 builds from this. G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948) International Covenant on Economic, Social and Cultural Rights, Dec. 16, 1966, 993 U.N.T.S. 3, *entered into force* Jan. 3, 1976 [hereinafter ICESCR].

199. *See generally* HUMAN GERMLINE GENOME MODIFICATION AND THE RIGHT TO SCIENCE (Andrea Boggio, Cesare Romano & Jessica Almqvist eds., 2020) (analyzing national regulatory frameworks around gene therapies and genome editing tools); *See also* Rumiana Yotova & Bartha Knoppers, *The Right to Benefit from Science and Its Implications for Genomic Data Sharing*, 31 EUR. J. INT'L L 665 (2020) (exploring the doctrine and normative content of right to benefit from science under Article 15 of the ICESCR); *see also* Rochelle Cooper Dreyfuss, *Human Rights in a Technological Age: The Right to Participate in Science*, 55 N.Y.U. J. INT'L & POL. 581 (2023) (arguing for a reinterpretation of the right to share in scientific advancement).

200. Econ. & Soc. Council, Comm. on Econ., Soc. & Cultural Rts., General Comment 25 (Sixty-seventh session) U.N. Doc E/C.12/GC/25 (addressing Article 15 of the ICESCR but largely completed before the COVID-19 pandemic refocused wide attention on virus and GSD issues).

V. Conclusion: Attributive and Distributive Justice in Genomic Sequence Data Sharing

This article has drawn attention to attributive justice as a concept which captures some important but elusive elements of human social practice which are influential in the construction and in the critique of legal and governance regimes. It has explored the place of this concept in claims and practices with regard to virus genomic sequence data. These claims are of two different types. One is made by people involved in the scientific process of collecting biological samples and sequencing and analyzing these. They seek recognition in subsequent publications made by others using this data, as well as some ability to set conditions on the use of this data, in addition to possibilities of research collaboration. This type of claim is now strongly embedded in scientific norms and in practices of grant funders and journals, and it is increasingly facilitated by digital sequence infrastructures. It has the attraction that fair attribution is a form of justice which resonates very widely across cultures. It has also a strong mobilizing effect because of its invocation as a means to counter colonial-type injustices and to leverage more equitable distributions of material resources and of status for scientific work. The worldwide benefits of this mean that this form of attributive justice is readily seen as aligned with global public and material interests.

A second type of claim is made by states (and with variations by indigenous peoples and communities) seeking to condition access to and use of sequences made from resources in a particular country or people or community, on a share of economic or health benefits attributable to those resources flowing back to that country or people or community. Yet there has been very little flow of benefits back to specific local communities or entire states under the Nagoya Protocol system for physical samples, or under cognate national legal arrangements. The dematerialized features of GSD and the complex ways in which GSD are actually used on pathways to commercial development make it unlikely that benefits will ever outweigh the aggregate transaction costs and deadweight

losses from extending to pathogen GSD the Nagoya Protocol's attributive approach.[201]

Leaders of developing countries and many civil society leaders within these countries have placed heightened emphasis on basic demands for compensatory or distributive justice in global negotiations.[202] The November 2022 agreement at the UN Framework Convention on Climate Change COP27 to establish a "Loss and Damage" Fund,[203] and the December 2022 commitment of developed countries to channel substantial resource transfers to developed countries for biodiversity purposes made as part of the Convention on Biodiversity post-2020 framework, are outcomes of an insistence on a form of compensatory justice for past over-use of what could have been a fairly shared and maintained commons.[204] The purely distributive agenda has met with less success, despite the rise and

---

201. A detailed menu of technical innovations to try to tie GSD to particular samples from particular locations, and to enable some eventual linkage of the GSD to commercial products, is developed in PAUL OLDHAM, DIGITAL SEQUENCE INFORMATION - TECHNICAL ASPECTS (2020). His report seeks to show that in aggregate a set of measures necessary to make an attributive ABS regime viable for GSD might be possible. But the complexity and intricacy of what must be put in place, suggest that this scenario is unlikely to be realizable in practice.

202. The UN Secretary-General put this perception starkly in January 2023: "the North-South divide is deepening. I am not convinced, that the wealthier world and their leaders truly grasps the degree of frustration and even anger in the Global South. Frustration and anger about the gross inequity of vaccine distribution in the recent past. Frustration and anger about pandemic recovery – with support overwhelmingly concentrated in wealthier countries that could print money. And trillions were printed in the global North, and of course developing countries could not print money because their currencies would go down the drain. Frustration and anger about a climate crisis that is crippling countries that contributed least to global heating. And the lack of the financial resources to respond to the challenge. Frustration and anger over a morally bankrupt financial system in which systemic inequalities are amplifying societal inequalities." U.N. Secretary-General, Remarks at the World Economic Forum (Jan. 18, 2023) https://www.un.org/sg/en/content/sg/statement/2023-01-18/secretary-generals-remarks-the-world-economic-forum [https://perma.cc/L6VQ-FQQ8].

203. Bharat Dahiya & Mahesti Okitasari, *Accessing the Loss and Damage Climate Fund,* UNITED NATIONS UNIVERSITY (2023), https://ourworld.unu.edu/en/accessing-the-loss-and-damage-climate-fund [https://perma.cc/AEL8-VGC4].

204. Conference of the Parties to the Convention on Biological Diversity (15th Meeting), *Kunming-Montreal Global Biodiversity Framework*, U.N. Doc. CBD/COP/15/L25 (Dec. 19, 2022).

rise of inequality. The 2023 BBNJ Agreement sits somewhere in between—it is not compensatory (for the most part), and while it preserves some rhetorical space for attributive and distributive approaches, the design of the regime aims more at conservation, capacity-building and enabling participation. In relation to health, the WHO has pursued a distributive justice agenda for some years under its "health for all" programs, and has ramped up specific efforts with regard to influenza and pathogens with pandemic potential from the early 2000s onward, particularly after Indonesia's 2006 demarche. As the question of GSD capacity-building and data-sharing grew in importance, it became a prominent case in struggles for attributive, compensatory, and distributive justice. In 2018, the WHO proposed a draft "code of conduct for open and timely sharing of pathogen genetic sequence data during outbreaks of infectious disease."[205] This would have allowed considerable scope for developing countries and their scientists to protect what they regarded as attributive interests.[206] By 2022 the

205. WHO Draft Code, *supra* note 103.

206. WHO Draft Code *supra* note 103, lines 104–130 ("Where data providers are not concerned about retention of ownership of the data, databases without data access agreements (such as GenBank, ENA and DDBJ) are the default option for sharing. For situations where data providers seek retention of ownership of their data, alternative models with data access agreements (such as GISAID9 in influenza) have been used to facilitate rapid sharing of GSD.

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

> 1) A database for hosting pathogen sequence data that meets best practice standards for administration and security
>
> 2) A database access agreement in the form of a simple clickthrough interface that allows 'providers' to share data and allows any genuine individual receiver party ('user') to access the data in alignment with relevant collaboration principles as outlined in the terms and conditions of the data access agreement, including provisions on ownership, intellectual property and access and benefit sharing. These conditions should include the following elements:
>
> - allow for ownership of the sequence data to reside with the provider of the data uploaded to the platform;
> - anyone using data accessed from the platform is required to acknowledge/credit/potentially co-author with the providers as appropriate;
> - (secondary) users should, as appropriate, propose and seek to collaborate with the data providers, including joint analysis of data;

WHO had shifted this emphasis somewhat in proclaiming the "WHO guiding principles for pathogen genome data sharing."[207] The overarching priority was rapid sharing of pathogen GSD to prevent, detect and respond to epidemics. This included local sequencing and analytical capacity building and geographical diversification. The INSDC and GISAID models were treated very evenly as being choices for the submitter, with each providing publicly accessible data and allowing for "global analyses and development of diagnostics, medicines and vaccines."[208] The conditional access models were in effect enjoined to do more to ensure conditions were actually complied with, including sanctioning breaches of the verified-user data access rules "as a means of supporting trustworthiness." The bigger focus of the WHO principles was now on specifically infrastructural elements: cross-platform collaboration on identifier systems and data reporting standards; aggregation to avoid partitioning of GSD datasets; large-scale standardized analyses; transparency and good governance of the infrastructures; sustainability of their financing mechanisms. Broader justice considerations were mentioned, but it was acknowledged that the infrastructural platforms were not themselves able to do a lot to achieve distributive justice.[209]

The WHO and CBD processes both involve a repeating dialogue and contestation between GSD infrastructural re-

---

• appropriate intellectual property management

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

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

207. WHO, *supra* note 14.

208. *Id.*

209. "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." *Id.*

gimes and their critics. They each focus serious attention on social, organizational, and technical dimensions of science that join together in these infrastructures. Changes in social, organizational, and technical practices interact with power fluxes and with swirling claims of justice to precipitate slow-moving adjustments and occasional abrupt equilibrium shifts.

Three broad points about an infrastructural approach to sequencing and virus justice, and a justice approach to GSD infrastructure, emerge from the discussion in this article.

First, the design and rules of the main GSD infrastructures play a key role—they are themselves a form of regulation, and they have come to be seen as increasingly important to the WHO and CBD regimes. The focus of WHO and CBD expert groups has become increasingly infrastructural, and this is now extending to data governance, including aspects of global data law, which are themselves means to regulate infrastructures.[210] Infrastructures and related data governance arrangements bear on topics ranging from openness and equity in scientific research and commercialization, to multilateral benefit-sharing in the BBNJ and plant regimes, to dual-use and national security concerns about data, artificial intelligence, and biology. They are likely to come into greater focus in all aspects of GSD law and governance. As GSD infrastructures are simultaneously technical, organizational, and social, changes in any of these dimensions may precipitate broader infrastructural change, though so far small-step changes have been the norm.

Second, attributive justice demands have commonalities, but work differently and have more or less epistemic and policy weight in different contexts. The GSD sharing regime has much in common with many other global governance regimes for quotidian cooperation, especially diffuse-source scientific cooperation. It blends private non-profit and for-profit endeav-

---

210. *See* Decision 15/9 on Digital Sequence Information, *supra* note 192 (taking a broad approach in acknowledging the FAIR and CARE (Collective benefit, Authority to control, Responsibility, and Ethics) data principles and their respective sub-principles); *see also* UNESCO, *UNESCO Recommendation on Open Science*, UNESDOC SC-PCB-SPP/2021/OS/UROS (Nov. 2021) and OECD, *Recommendation on Enhancing Access to and Sharing of Data*, OECD/LEGAL/0463 (Oct. 2021). Sagely in the light of all of this, the COP15 Decision assigned principles of data governance as an issue requiring further consideration.

ors with those of state institutions, informal science transnationalism with formal inter-governmentalism, and richer countries with poorer countries—but the overall balance has been skewed to favor the preferences of rich countries. It steers between the demands for a well-designed and limited-cost functional regime with enough support to work effectively within prevailing scientific-political constraints, and a range of acute or chronic demands for a regime which meets basic requirements of justice. In steering that path, it has tended to prioritize the functional over all but the most functionally necessary revisionist demands of justice, guided by what is thought internally to be a sage disposition to avoid "politicization" and to maintain buoyancy by not taking on too much weighty baggage. While this functionally-focused equilibrium was sustainable for several decades, it has come under pressure from rising discontent with structural inequality and disregard, and has been juddered by technological change, demographic shifts, and epidemic catastrophes. These have precipitated an intensification of claims which can be understood as demands for justice. The clearest of these has been for attributive justice. Claims for distributive justice or compensatory justice have been thwarted, and often submerged, because of the evident limitations of the intermediating GSD infrastructures or of frontline scientists or health agencies as agents capable of delivering distributive or compensatory justice at scale. Attributive justice has been a strong basis of demands for scientific recognition—these have been facilitated by the existence of GISAID as a significant alternative to GenBank and the INSDC. To some extent, these demands operate as a counterpower, in which voices critical of existing power distributions and structures have received some amplification. Demands by states for attributive justice over virus GSD, articulated as "virus sovereignty" or "genomic sovereignty" and as extension to GSD of benefit-sharing if not of the access-benefit structure of the NP, are supported by most developing countries. But there is significant resistance in health research and policy, including from the WHO as well as from scientists working with INSDC and GISAID, to simply extending arrangements about material samples of pathogenic viruses to GSD, on the grounds that this impairs urgent health research and medical response and is anyhow inefficient and ineffective. There also remain concerns that the NP idea of location-specific rents, extended

to pandemic risks, has the morally contestable effect that greater risk generates greater leverage for side-payments ("benefits"). Rapid developments in synthetic biology and in AI, with models trained on GSD from actual samples but also on synthetic data, will complicate attribution and heighten inequities. Hybrid distributive-attributive justice forms are likely to develop in relation to epidemics, with the PIP as one model. Little progress can be expected in normative attempts to tightly condition flows of GSD as an exchange for flows to specific sample-originating territories of other benefits such as diagnostics, therapies, and vaccines that arise downstream.[211] Infrastructural arrangements, conditions for, and controls on data and data flows may however have significant effects in practice toward modifying this equilibrium. The blending of attributive, compensatory, and distributive justice objectives, in the era of digital sequencing as a major part of combatting virus epidemics, must in some measure be pursued through the construction, adaptation, repurposing, and governance of infrastructures.

Third, governance transformations, including projects of democratization and contestation, entail challenging the "cartography that maps questions of fact as the domain of experts and questions of value as the domain of democratic choice."[212] Asserting and making truly operational the publicness of infrastructures—their responsibilities to speak and act for, and with regard to all of the publics on whom they have impact[213]—is a necessary normative and functional dimension of sequence and virus justice, and of globally effective management of pathogenic threats. The instantiation of attributive justice in the GSD regime, viewed in Grotian terms, contributes to an order of respect and recognition which may become even more important to the *salus populi* and shared future existence.

---

211. Bagley, *supra* note 153, at 53–61.

212. Stephen Hilgartner, Clark Miller & Rob Hagendijk, *Introduction, in* SCIENCE AND DEMOCRACY: MAKING KNOWLEDGE AND MAKING POWER IN THE BIOSCIENCES AND BEYOND 2, 3 (Stephen Hilgartner, Clark Miller & Rob Hagendijk eds., 2015).

213. Benedict Kingsbury & Nahuel Maisley, *Infrastructures and Laws: Publics and Publicness*, 17 ANN. REV. L. & SOC. SCI. 353 (2021).