A Comparative Review of Data Sharing Regulations and Practices in Four Jurisdictions (Australia, France, United Kingdom, United States)

Sina Edwin Faraji, Hanshi Liu, Aurelie Dauge, Beatrice Kaiser, Palmira Granados-Moreno & Yann Joly

Centre of Genomics and Policy
McGill University
740, avenue Dr. Penfield, suite 5200
Montreal (Quebec)
H3A 0G1
CANADA

Report Submitted to Genome Canada on 2022-04-04 in the context of the CanCOGeN project

CGP
CENTRE OF GENOMICS AND POLICY
CENTRE DE GÉNOMIQUE ET POLITIQUES

Abstract:

This report compares the legal and organisational differences of four countries – the US, the UK, France and Australia – with regard to genetic and health data sharing initiatives both domestically and internationally. Reviewing statistics on the pace of viral sequence data deposits to international databases such as GISAID, this paper explains why some countries may achieve more complete and quicker health data sharing. Although performance in sharing is more easily measured for viral sequence data due to the challenge in obtaining similar statistics for human genomic data in controlled access, measuring performance remains relevant for host sequence data as well. As such, conclusions are also provided for more sensitive data such as host sequence data or viral sequence data with extensive metadata. For more sensitive information, we noted some legal and structural differences such as stricter formalities or the presence of more than one consortium that could impact the pace at which data is shared. For less sensitive data such as viral sequence data with minimal metadata, our comparison is inconclusive. Interestingly, while all the countries analyzed had very similar legal restrictions and analogous organisational structures, they performed very differently at sharing their genomic and metadata.

INTRODUCTION

The COVID-19 pandemic has increased the need for accurate and timely access to information, especially for public health research that is essential to allow governments to take rapid measures to address COVID-19 variants. In April 2020, Genome Canada launched the Canadian COVID-19 Genomics Network (CanCOGeN), an initiative composed of two large projects. On one hand, VirusSeq is responsible for sequencing viral genomes and sharing them with accompanying metadata to the community and, on another, HostSeq focuses on host genome and metadata. Both projects aim to facilitate and encourage sharing and linkages of viral and patient data for public health research. Similar data-sharing efforts have been set up in countries all over the world.

The *Global Initiative on Sharing All Influenza Data* (GISAID) international initiative encourages sharing viral sequence genetic information related to COVID-19 by providing a repository for the data and overseeing access. GISAID is a registered access, free-of-charge, database. The *Access agreement* [1] used by GISAID does impose some limitations on data usage. Similarly, the *European Nucleotide Archive* (ENA) is a repository that specifically collates sequence data. It includes a comprehensive list of information such as raw sequencing data, sequence assembly information and functional annotation. While these repositories allow for better data communication internationally, they only host de-identified data to simplify the requirements of the sharing process. There is currently no open access centralized international platform for sharing more sensitive genetic information such as host sequence data.

In Canada, the sharing of data for both VirusSeq and HostSeq started at a slow pace at the beginning of the pandemic; early in 2021, things started to gradually pick up.[2] A similar phenomenon is observable globally where analogous initiatives are put into place. The variation of deposition times of unidentifiable virus sequence data into GISAID per country[3] poses an important question: What barriers to viral genetic sequence data sharing can explain the differing performance of various jurisdictions having contributed to this database? Although performance in sharing (pace of deposits and completeness of dataset) is more easily measured for viral sequence data in GISAID due to the challenge in accessing similar statistics for human genomic data[4], measuring performance remains relevant for HostSeq. This report will look at both

---

[1] "GISAID - Terms of Use", online: GISAID <https://www.gisaid.org/registration/terms-of-use/>.

[2] CanCOGeN VirusSeq Data Sharing Committee, "CanCOGeN Pan-Canadian Experience In Data Sharing: Lessons Learned and Recommendations for the Road Ahead", Report [forthcoming in 2022].

[3] See Appendix 2A.

[4] See Appendix 2A.

the legal and organisational/practical context around the sharing of genomic data and accompanying metadata in the United States of America (US), the United Kingdom (UK), France and Australia. Specifically, three types of information will be considered: *(1) viral sequence data including limited non-identifying metadata, (2) viral sequence data including extensive metadata, and (3) host sequence data with extensive metadata.* While (2) and (3) may allow for the identification of hosts, information of type (1) most likely does not allow for any personal identification.[5] The comparison between the aforementioned countries will allow for the better identification of legal and structural barriers to data sharing globally and may help uncover a path forward for Canadian genome data sharing for public health and research.

I. AUSTRALIA

A. Legal Framework regulating data sharing

The federal legal framework regulating the collection, use, and disclosure of personal data in Australia is mainly defined by the *Privacy Act 1988*.[6] The *Privacy Act* applies to private health service providers, but not to state and territory public sector health service providers (e.g., hospitals) and to state or territory government agencies for certain healthcare acts.[7] State and territory public hospitals or health care facilities are covered under specific state and territory privacy legislation if the state has enacted such legislation. In such states, these laws will also impose requirements on private-sector health services.[8]

The *Privacy Act* defines *personal information* as "information or an opinion about an identified individual, or an individual who is reasonably identifiable"[9] and states that within what is considered personal information, health, genetic and biometric information about an individual is deemed "sensitive

---

[5] Song et al, "Addressing Privacy Concerns in Sharing Viral Sequences and Minimum Contextual Data in a Public Repository During the COVID-19 Pandemic" (2022) 12 Frontiers in Genetics, online: <https://www.frontiersin.org/article/10.3389/fgene.2021.716541>.

[6] *Privacy Act 1988* (Cth), 1988/119.

[7] "Rights and responsibilities", online: Office of the Australian Information Commissioner (OAIC) <https://www.oaic.gov.au/privacy/the-privacy-act/rights-and-responsibilities>.

[8] Private sector health service providers must comply with federal and state privacy laws in New South Wales, Victoria, and Australian Capital Territory; only public health service providers must comply with privacy legislation in Queensland, Tasmania, and the Northern Territory; South Australia and Western Australia don't have specific privacy legislation and privacy issues are handled by specific agencies such as privacy committees. See "Privacy in your state", online: OAIC <https://www.oaic.gov.au/privacy/privacy-in-your-state>.

[9] *Supra* note 6 at s 6.

information".[10] Sensitive health information may be disclosed in certain defined circumstances such as "permitted health situations" relating to research, genetic information, and cases of a responsible person for an individual.[11] When it is impractical for researchers to obtain consent, researchers are allowed a consent waiver to use personal information for health and medical research purposes, including genetic information, in specific situations.[12] The National Health and Medical Research Council (NHMRC), which oversees the human research ethics framework in Australia, has also implemented legally binding guidelines for the use of health and genetic data without consent.[13] While it may not be required for the collection, use or disclosure of sensitive information, consent may still be required to participate in a research study, according to the National Statement on Ethical Conduct in Human Research (NSECHR).[14]

The Office of the Australian Information Commissioner (OAIC) is responsible for enforcing data protection under the *Privacy Act*.[15] In accordance with this regulatory framework, collection of sensitive information, such as health and human genetic information, requires consent, express or implied,[16] and use should be limited to cases related to the specific purpose for which consent was granted.[17] It is expected that reasonable measures are employed to protect this information from "misuse, interference and loss, and unauthorized access, modification or disclosure".[18] These requirements are in line with the *National COVID-19 Privacy Principles*, created as a consistent national framework on handling personal information, which stipulate that personal information be stored in Australia and state an expectation that information provided to maintain the COVID-19 response be destroyed when no longer needed.[19]

In the context of the COVID-19 pandemic, all privacy regulators in Australia (federal, state, and territory) have formed a National COVID-19 Privacy Team and created the universal *National COVID-19 Privacy*

---

[10] *Ibid*.

[11] *Ibid* at Schedule 1 Part 2 s.3.3 and 3.4 and Part III Division 2 s.16B.

[12] *Ibid* at s 95, 95A and 95AA.

[13] *Ibid* at Part IX s 95, 95A, 95AA; "Health and medical research", online: OAIC <https://www.oaic.gov.au/privacy/the-privacy-act/health-and-medical-research>.

[14] See *National Statement on Ethical Conduct in Human Research*, 1864962755 2007. s 2.3.10.

[15] "What we do", online: OAIC <https://www.oaic.gov.au/about-us/what-we-do>.

[16] *Supra* note 6 at Division 1 s 6 and Schedule 1 Part 2 s 3.3 and 3.4.

[17] *Ibid* at Schedule 1 Part 3 s 6.1 and 6.2; "National COVID-19 Privacy Principles" (2 September 2021), online: OAIC <https://www.oaic.gov.au/privacy/guidance-and-advice/national-covid-19-privacy-principles>.

[18] OAIC, *supra* note 17; *supra* note 6 at Schedule 1 Part 4 s 11.

[19] OAIC, *supra* note 17.

Principles.[20] Recently, regulations were also implemented in the *Privacy Act* to dictate privacy measures relating to personal data collected via Australia's COVID app data.[21] The Australian Government Department of Health (Dept. Health) sets out how they manage personal information under the *Privacy Act* in their *Privacy Policy*.[22] In cases of public health emergencies, Australia adopted the *National Health Security Agreement 2011*, which supports their *National Health Security Act 2007*.[23] It promotes information sharing and allows personal information disclosure for a coordinated national response to public health crises, though genetic and viral data sharing are not explicitly mentioned.[24]

B. Concrete actors and practices

1. COVID-19 health data

Australia's Department of Health publicly releases bi-weekly epidemiology reports on COVID-19 issued by the *Communicable Diseases Intelligence* (CDI) Coronavirus disease (COVID-19), using data from health surveillance systems, including the *National Interoperable Notifiable Disease Surveillance System* (NINDSS)[25] and the *National Incident Room* (NIR) to assess the severity of COVID-19.[26] Requests to access the specific data behind these reports are currently on hold and would require agreements with the states and territories via the *Communicable Diseases Network Australia* and may require ethics approval.[27]

2. COVID-19 viral sequence data with minimal metadata

---

[20] "COVID-19 response from Australian privacy regulators" (27 March 2020) online: OAIC <https://www.oaic.gov.au/updates/news-and-media/covid-19-response-from-australian-privacy-regulators>; OAIC, *supra* note 17.

[21] *Supra* note 6 at Part VIIIA.

[22] Austl, Commonwealth, Department of Health, *Privacy Policy* (October 2020).

[23] Austl, Commonwealth, *National Health Security Agreement* (July 2011).

[24] *Ibid*.

[25] The NINDSS is in the process of replacing the previous system, the NNDSS.

[26] "Coronavirus disease (COVID-19) epidemiology reports, Australia, 2020–2021" (8 November 2021), online: Department of Health <https://www1.health.gov.au/internet/main/publishing.nsf/Content/novel_coronavirus_2019_ncov_weekly_epidemiology_reports_australia_2020.htm>; Austl, Commonwealth, Communicable Diseases Intelligence, *COVID-19 Australia Epidemiology Report* by COVID-19 National Incident Room Surveillance Team, Communicable Diseases Intelligence: Technical Supplement (Australian Government Department of Health, 2021).

[27] Department of Health, *supra* note 26.

AusTrakka is Australia's national platform established by the *Communicable Diseases Genomics Network* (CDGN) that allows for Australia (and New Zealand)-wide real-time surveillance and analysis of pathogen genomics.[28] In the context of COVID-19 viral data, AusTrakka has allowed all public health laboratories across Australia and New Zealand to access AusTrakka and upload their SARS-CoV-2 sequences to allow surveillance of cases and variants nationally.[29] Data from CDGN laboratories is shared with the NIR and included in CDI COVID-19 epidemiology reports.[30] Data sharing practices are outlined in a non-legally binding framework.[31] Requests to access AusTrakka compiled data for research must be approved by the CDGN.[32] The public health labs retain custodianship of their sequence and metadata.[33]Metadata collected for AusTrakka, at minimum, includes laboratory, patient jurisdiction and date of collection, with other recommended information.[34] Public health labs that contribute to AusTrakka must be individually requested to share their own sequence, meta, and epidemiological exposure data.[35] Public health labs can independently share viral sequence data and non-identifying metadata with international databases such as GISAID.[36] At the international level, access to AusTrakka data by any country (except for New Zealand) requires an approved request, which is governed by the *National Health Security Act* and *National Health Security Agreement*.[37] Access requests for research and development purposes may be submitted, and if successful will require a contract or research agreement.[38]

3. COVID-19 host sequence data with metadata and viral sequence data with extensive metadata

Although Australia does not have a nationally organized effort to share host genomic data associated with viral genomic data, given that Public Health Labs (PHLs) have decision-making power over their sequenced data, they can decide to independently contribute to international projects. In fact, several Australian labs

---

[28] "National genomics reporting", online: Communicable Diseases Genomics Network (CDGN) <https://www.cdgn.org.au/national-genomics-reporting>; "AusTrakka: Overview", online: CDGN <https://www.cdgn.org.au/austrakka>; and AusTrakka & Communicable Diseases Genomics Network, *Framework for data sharing and analysis for SARS-CoV-2 in the AusTrakka system* (2020).

[29] *Ibid*.

[30] CDGN, *supra* note 28.

[31] AusTrakka & Communicable Diseases Genomics Network, *supra* note 28.

[32] *Ibid*.

[33] *Ibid*.

[34] Recommended metadata include patient age, gender, date of symptom onset, sequencing technology and strategy, and optional epidemiological exposure data; *ibid*.

[35] *Ibid*.

[36] *Ibid*.

[37] *Ibid*.

[38] *Ibid*.

are involved in the "COVID Human Genetic Effort", an international consortium investigating human genetic and immunological bases of COVID-19 infection.[39] The data shared with this consortium is not publicly available but is shared on a common hub between participating centres internationally.[40] As for viral sequence data, it should be noted that the amount of metadata shared with platforms such as AusTrakka remains at the discretion of PHLs. Hence, in addition to the minimal metadata stored by AusTrakka, the platform may also be hosting more sensitive data.

II. UNITED STATES

A. Legal Framework regulating data sharing

The US has adopted privacy and privacy-related legislation both at the federal and state levels.[41] Concerning genomic information and data sharing, the *Health Information Portability and Accountability Act* (HIPAA) regulates, at the federal level, the protection of personal information collected and held by "covered entities"[42] in the course of health care services.[43] While the HIPAA *Security Rule* (the "Security Rule") defines requirements for keeping individually identifiable "protected health information" (PHI) secure, the HIPAA Privacy Rule (the "Privacy Rule") governs the collection and disclosure of HIPAA-covered individually identifiable information.[44]

The Security Rule implements various security requirements by using appropriate administrative, physical, and technical safeguards to ensure the confidentiality, integrity, and security of information. The Security Rule applies to all individually identifiable health information a covered entity creates, receives, maintains, or transmits in electronic form.[45] The Privacy Rule is the primary and most consistently applicable standard in the US privacy protection landscape. In fact, it has predominance over any state law that is contrary to

---

[39] "Our Mission", online: COVID Human Genetic Effort (CHGE) <https://www.covidhge.com/>.

[40] "Sequencing Hub Leaders", online: CHGE <https://www.covidhge.com/sequencing-hubs>; "Participating Centers Regional Coordinators", online: CHGE <https://www.covidhge.com/participating-centers>.

[41] US, Pittman & Levenberg (Global Legal Group), *International Comparative Legal Guides*; US Department of Health and Human Services, *Summary of the HIPAA Privacy Rule* (2013), <(https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html> Last modified: 27 July 2021).

[42] HIPPA Rules only apply to "covered entities and business associates". Covered Entities in general include Health Care Providers such as doctors, pharmacies, nursing homes, clinics, insurance companies, certain government healthcare programs, and medical claims organizations, Business associates refer to the entities that covered entities engage in the course of fulfilling their HIPAA based activities.

[43] US Department of Health and Human Services, *supra* note 41, s Introduction.

[44] *Ibid*.

[45] *Ibid.*

it. The Privacy Rule focuses on the protection of "individually identifiable health information".[46] That is, data that can be reasonably attributed to an identifiable individual, including data that relates to the past, present, and future physical or mental health of an individual and the provision of healthcare to an individual. [47] To the extent that it allows a person to be individually identified, information specific to COVID-19 would also be covered by the Privacy Rule, especially given that laboratories and other healthcare settings are considered "covered entities" under the law.[48] In the context of COVID-19, properly de-hosted viral sequence data with minimal metadata would generally not be considered protected health information given that it should not allow for the identification of an individual. The same cannot be said, however, about host sequence data or viral sequence data linked with extensive metadata.

Protected health information can only be used and disclosed when permitted by HIPAA, which, in most cases, will require the consent of the individual source of the data.[49] However, HIPAA provides for exceptional circumstances where consent from the individual is not required, some of which are pertinent to the COVID-19 context. For instance, consent is not required from the individual for uses and disclosures related to public health activities, health oversight activities, research and to counter serious threats to health and safety.[50] It should be noted, however, that even though consent is not required for the research use or disclosure of protected health information, consent may still be required to participate in a research study, according to the Common Rule.[51] Some exceptions may apply under the Common Rule, where the requirement for consent could be waived or even fall under an exempt situation.[52] In general, despite the existence of these exceptions, such situations are still contingent on the researchers meeting various conditions.[53]

B. Concrete actors and practices

---

[46] *Ibid* at s State Law.

[47] *Ibid* at s Introduction.

[48] *Ibid* at s Limiting Uses and Disclosures to Minimum Necessary.

[49] *Ibid* at s General Principles for Uses and Disclosures.

[50] *Ibid* at s Public Health Activities.

[51] The Common Rule is the short name for the Federal Policy for the Protection of Human Subjects and applies to human subject research. It is a standard of ethics by which all government-funded U.S. research is held to. Among other things, it describes the type of research subject to regulation, provides definitions, sets forth requirements for an Institutional Review Board's authority and states requirements for informed consent.

[52] Office for Human Research Protections (OHRP), "2018 Requirements (2018 Common Rule)" (7 March 2017), online: HHS.gov <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html> (Last modified: 06 August 2021 14:38:48-0400).

[53] US Department of Health and Human Services, *supra* note 41 at s Permitted Uses and Disclosures; OHRP, *supra* note 47 at s General Requirements for Informed Consent.

1. *COVID-19 viral sequence data with minimal metadata*

Regarding viral sequence data with minimal metadata, the United States Center for Disease Control's (CDC) SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology and Surveillance (SPHERES) initiative is the primary facilitator of COVID-19 related genomic data sharing in the United States.[54] SPHERES is a continuation of the CDC's Advanced Molecular Detection program, which for the past seven years has centered around expanding the use of pathogen genomics, bioinformatics, and epidemiology to boost infectious disease surveillance and outbreak responses.[55] It consists of an active collaboration between over 60 federal agencies, federal laboratories, state/local laboratories, academic institutions, corporations, non-profit organizations, and international collaborators assembled to coordinate and standardize COVID-19 sequencing across the United States.[56] SPHERES does not collect data from COVID-19 patients. Rather, labs and other data providers have the option of submitting their data to public databases through SPHERES, or they may provide their data to the CDC, which then submits the data on their behalf.[57] Submitting such data through SPHERES may be practical for labs that lack the capacity to directly submit to public repositories themselves.

The National Center for Biotechnology Information (NCBI) receives data from data contributors and provides access to genomic information to the international community. [58] NCBI relies on contributors such as individual laboratories and large-scale sequencing projects to receive SARS-CoV-2 sequences.[59] Submissions are made through the NCBI GenBank, responsible for database collection of all publicly available nucleotide sequences and their protein translations.[60] Contributors can use the web-based tools for automatic submission to GenBank or they can submit the data via the GenBank Submission Portal or email.[61] GenBank does not store personal information and advises contributors to not include any data that

---

[54] "SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES)" (11 February 2020), online: Centers for Disease Control and Prevention (CDC) <https://www.cdc.gov/coronavirus/2019-ncov/variants/spheres.html>.

[55] *Ibid* at s SPHERES Overview.

[56] *Ibid.*

[57] "CDC's Role in Tracking Variants" (8 September 2021), online: CDC <https://www.cdc.gov/coronavirus/2019-ncov/variants/cdc-role-surveillance.html>.

[58] "GenBank Overview", online: NCBI <https://www.ncbi.nlm.nih.gov/genbank/>.

[59] Dennis A Benson et al, "GenBank" (2002) 30 Nucleic Acids Research 1 at 17–20.

[60] *Ibid*; *supra* note 58.

[61] "How to submit data to GenBank", NCBI online: <https://www.ncbi.nlm.nih.gov/genbank/submit/>.

could reveal the personal identity of the source of the data when submitting human sequences.[62] The data repository also provides public access to the viral sequences, notably through the NCBI Virus SARS-CoV-2 Data Hub Interactive Dashboard or, simply, by downloading viral genomes and protein sequences, annotation, and data reports from NCBI Datasets.[63]

2. *COVID-19 host sequence data with/without metadata and viral sequence data with extensive metadata*

While viral sequence data with limited metadata for COVID-19 is largely kept by the NCBI in the U.S., some of the more sensitive data is made available by the NCATS National COVID Cohort Collaborative (N3C) Data Enclave.[64] The N3C is a part of the larger NIH National Center for Advancing Translational Sciences (NCATS) and is used to study COVID-19 as well as its risks and health consequences.[65] The N3C directly collects the information case-by-case through collaborating medical institutions and health care organizations that make their data available to the N3C effort.[66] The N3C team must ensure that the submitted data meets legal and regulatory requirements, such as the Privacy Rule, before storing it.[67] Data sets received by the N3C include extensive meta information such as demographics, symptoms, lab test results, procedures, medications, medical conditions, physical measurements and more, but no host sequence data.[68] Access to data stored within the repository is provided based on a tiered system. In order from more to less restrictive[69], the data is categorized into 3 groups – Limited Data Set (LDS), De-identified Data Set (DDS) and Synthetic Data Set (SDS) – and each group has a defined set of conditions that must

---

[62] *Ibid* at s Confidentiality.

[63] "NCBI SARS-CoV-2 Resources", s Explore the Data, online: NCBI <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

[64] "About the National COVID Cohort Collaborative" (1 February 2022), online: National Center for Advancing Translational Sciences <https://ncats.nih.gov/n3c/about>.

[65] *Ibid*.

[66] *Ibid* at s What data does the N3C have and where does it come from?.

[67] *Ibid* at s How does the N3C keep data secure and protect patient privacy?.

[68] *Supra* note 64.

[69] NIH, "N3C Data Overview" (31 August 2020), online: National Center for Advancing Translational Sciences <https://ncats.nih.gov/n3c/about/data-overview>. at s Access Requirements for Researchers by Data Level.
From the NIH website on the data provided https://ncats.nih.gov/n3c/about/data-overview: (1) Limited Data Set (LDS): Consists of patient data that retain the following protected health information — dates of service; patient ZIP code. (2) De-identified Data Set: Consists of patient data from the LDS with the following changes — Dates of service are algorithmically shifted to protect patient privacy; patient ZIP codes are truncated to the first three digits or removed entirely if the ZIP code represents fewer than 20,000 individuals or represents Tribal lands. (3) Synthetic Data Set: Consists of data that are computationally derived from the LDS and that resemble patient information statistically but are not actual patient data.

be met for access.[70] For each tier of access, the patient described is the same. However, the metadata descriptors available are different.[71] Given that it is more restrictive, the LDS dataset uses fewer descriptors than the DDS dataset.[72] Finally, for the SDS dataset, the metadata descriptors available are computationally generated to resemble the LDS set but are not the actual LDS data.[73] Additionally, only researchers from US-based institutions can access LDS.[74]

III. UNITED KINGDOM

A. Legal framework regulating data sharing

The UK legal framework that protects personal data is composed mainly of two laws: the *UK General Data Protection Regulation* (UK GDPR) and the *Data Protection Act of 2018* (DPA 2018).[75] Both laws apply across different sectors and industries, such as healthcare, the commercial sector, or the public sector regardless of the way each member of these sectors "processes" "personal data".[76] Consequently, data that is not deemed to be "personal" or that is not "processed" does not fall within the aforementioned laws. Thus, viral data with minimal metadata that does not allow for the reidentification of individuals is likely exempt from them. On the other hand, host sequence data or viral sequence data with extensive metadata, if processed, would fall within their scope. In fact, the UK GDPR creates a "special category" for genetic data or data concerning health[77] that can only be processed under specific conditions when necessary for

---

[70] *Ibid*.

[71] *Ibid*.

[72] *Ibid*.

[73] N3C registration, N3C Data Enclave account, Data Use Agreement (DUA) executed with NCATS, NIH IT training completion, Approved Data Use Request (DUR), Human Subjects Research Protection training completion and Local Human Research Protection Program IRB determination letter.

[74] *Ibid*.

[75] "Guide to the UK General Data Protection Regulation (UK GDPR)" (29 March 2022), online: UK ICO <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>.; UK Parliament, *(Retained EU Legislation) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation* (UK GDPR).

[76] UK Parliament, *supra* note 75 at 4*;* Article 4(1) of the UK GDPR defines "personal data" as "any information relating to an identifiable or identified individual". Article 4(2) defines processing as any operation or set of operations, which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

[77] UK Parliament, *supra* note 75 at 9 (1).

reasons of substantial public interest or for public interest in the area of public health.[78] Special category data that is subject to article 9 of the UK GDPR is also subject, more broadly, to article 6 which describes lawful processing.[79] Although both articles provide for lawful processing when consent is obtained, article 9's consent threshold is more rigorous as it requires "explicit consent".[80] Accordingly, it should be noted, however, that consent is not necessarily required to process host sequence data or viral sequence data with extensive metadata.[81] For example, host sequence data that is used or disclosed while necessary both for the performance of a task carried out in the public interest and for reasons of public interest in the area of public health will generally not require consent.[82] Research aimed at addressing the COVID-19 pandemic would most likely fall within such a situation.[83]

This framework also regulates the transfer and sharing of data to countries outside of the UK. Other countries must provide an adequate level of protection that is similar to that in the UK[84] and provide appropriate data safeguards.[85]

B. Concrete actors and practices

In the UK, the principal consortium involved in sharing SARS-CoV-2 genomic data and associated contextual data is the COG-UK consortium. COG-UK is a partnership composed of several academic institutional partners, diagnostic laboratories, and National Health Service (NHS) laboratories.[86] The consortium collects the data and provides access to SARs-CoV-2 genomic data sharing to facilitate efforts

---

[78] UK Parliament, *supra* note 75 at 9 (2) g) and i).

[79] UK Parliament, *supra* note 75 at 6 and art.9.

[80] UK Parliament, *supra* note 75 at 9.

[81] UK Parliament, *supra* note 75.

[82] Application of article 6 (1) e) and article 9 (2) i) of the UK GDPR. Article 9 (2) i) does require that it is processed on the basis of the law of the United Kingdom or a part of the United Kingdom.

[83] Regina Becker et al, "COVID-19 Research: Navigating the European General Data Protection Regulation" (2020) 22 Journal of Medical Internet Research 8, e19799. https://doi.org/10.2196/19799.

[84] "International transfers after the UK exit from the EU Implementation Period" (31 January 2022), online: UK ICO <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/international-transfers-after-uk-exit/>. Under the UK GDPR and DPA 2018, certain countries are considered to hold privacy protection laws that provide a similarly adequate level of protection. Data sharing to these countries can therefore be based on "adequacy regulations".

[85] Ibid at s Is the restricted transfer covered by appropriate safeguards?. Under the UK GDPR and DPA 2018, a transfer of personal data based on appropriate safeguards such as a legal instrument ensuring appropriate safeguards for the protection of personal data can be used. This is according to Chapter 5 Article 75 of the DPA 2018.

[86] "About Us | COVID-19 Genomics UK Consortium" (14 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/about/about-us/about-us/>.

in research and public health to address the pandemic.[87]COG-UK stores both viral sequence data with extensive or limited metadata within CLIMB-COVID, a secure digital cloud infrastructure that serves as a database.[88]

The COG-UK model proposes three levels of data with specific terms and conditions of data access for each level. The three levels proposed are Public, Consortium, and Restricted, with each advancing category of data being more descriptive, more sensitive, and more identifying.[89] The public data level contains only non-identifying low-level contextual data and the associated genomic sequence which is the case for a viral genomic sequence with minimal metadata.[90] This level has no official publication policy but acknowledging those who have generated the data and offering them the possibility of getting involved in the data analysis is encouraged.[91] The Consortium level comprises more detailed information such as the age, sex, and swab site (location of the lab where the sample was collected).[92] Since this level offers access to the public data in addition to the above-mentioned more detailed information, access requires COG-UK consortium membership.[93] This membership requires users to apply to the COG-UK consortium and to sign the COG-UK consortium agreement.[94] The restricted level provides access to the most sensitive and identifying data fields of data, such as individual ethnicity and socioeconomic status.[95] In addition to the requirements of consortium-level data access, the restricted level of data access requires a specific data-access agreement (DAA).[96] This DAA is evaluated by the COG-UK steering group and data-access representatives from each of the four involved public health agencies (England, Northern Ireland, Scotland, and Wales) whose data

---

[87] *Ibid.*

[88] "CLIMB | COVID-19 Genomics UK Consortium" (2 February 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/priority-areas/data-linkage-analysis/climb/>.

[89]Ewan Harrison, *Data Sharing Inside and Outside of the COVID-19 Genomics UK Consortium* (2021).

[90] *Ibid*.

[91] "Public Data & Analysis | COVID-19 Genomics UK Consortium" (12 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/>.

[92] "A short history of the COVID-19 Genomics UK (COG-UK) Consortium" (17 December 2020), online: COVID-19 Genomics (COG-UK) Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/a-short-history-of-the-covid-19-genomics-uk-cog-uk-consortium/>; "Reflections on the achievements of COG-UK | COVID-19 Genomics UK Consortium" (22 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/reflections-on-the-achievements-of-cog-uk/>; "Original structure | COVID-19 Genomics UK Consortium" (14 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/about/about-us/original-structure/>.

[93] *Supra* note 89.

[94] COG-UK Steering Group, *COG-UK Publication and Authorship Policy* (2021).

[95] *Supra* note 89.

[96] *Ibid.*

are involved.[97] In the case of Restricted data, both the COG-UK Steering Committee and PHA representatives have veto powers to reject such data access.[98] If access for restricted data is approved, a six-month data-access period will generally be allowed for.[99] For publications at the Consortium or Restricted levels, the involved PHAs and Steering Committee must be consulted. They have the discretion to veto any such publication.[100]

Outside of COG-UK, the UK Health Security Agency (UKHSA) and UK Department of Health and Social Care (DHSC) also play a role in facilitating access to protected data (such as viral data with more extensive contextual/metadata).[101] Both entities collaborate with industry, academia and other governmental organizations in generating, curating, and integrating data on infectious diseases and external public health threats, with UKHSA serving as the *data steward*, while the DHSC is the *data controller*.[102] Researchers based outside of the UK conducting public health research may access higher level data/protected data through the UKHSA.[103] This protected data can include COVID-19 related data, such as viral genomic sequences with extensive metadata.[104] This data is available through controlled access procedures. External researchers must submit a data access request detailing their proposed research project, the data necessary, project aims/objectives, and more.[105] Afterwards, this data access request will be formally evaluated by a Data Access Committee.[106] The UKSHA may then provide feedback, request further details, and impose conditions via a legally binding data-sharing contract.[107]This contract may govern conditions such as security requirements, data minimization, data retention, data reduction, confidentiality requirements and more.[108]

IV. FRANCE

---

97 *Ibid*.

98 *Ibid*.

99 Once approval is given, it is effective for 6 months. Beyond that period, access will have to be renewed.

100 *Ibid.*

101 "Accessing UKHSA Protected Data" (14 March 2022), online: UK Health Security Agency (UKHSA) <https://www.gov.uk/government/publications/accessing-ukhsa-protected-data/accessing-ukhsa-protected-data>.

102 *Ibid* at s Contents.

103 *Ibid*.

104 *Ibid* at s Improving Access to Data.

105 *Ibid* at s Applying to Access Protected Data.

106 *Ibid* at s Application Process Summary.

107 *Ibid*.

108 *Ibid*.

A. Legal Framework regulating data sharing

France has two levels of privacy-related laws – domestic and European – both of which cover only personally identifiable information.[109] At the European level, the *General Data Protection Regulation* (EU GDPR) aims to harmonize data protection laws across the European Union Member States.[110] Under the EU GDPR, Member States are allowed some flexibility and may maintain or introduce further conditions of the law, notably those applicable to the processing of genetic data, biometric data or data concerning health.[111] As such, at the domestic level, while the *French Data Protection Act 1978* (FDPA), which was amended following the adoption of the EU GDPR in 2018, incorporates provisions of the EU GDPR into French law,[112] it also adds a specific section on the processing of health-related data.[113]

Just as it is the case under the EU GDPR, the FDPA legitimizes the processing of personal health data without requiring consent when it is necessary both for the performance of a task carried out in the public interest and for reasons of public interest in the area of public health.[114] Just like for the UK GDPR, research aimed at addressing the COVID-19 pandemic would most likely fall within such a situation.[115] While consent may not be necessary under certain circumstances, the *French Public Health Code* (FPHC) makes use of consent a condition to undertaking research involving humans.[116] If used in research, host sequence data is subject to these provisions, despite not requiring the individual data source's consent when processed.[117]

---

[109] *Loi n°78-17 du 6 janvier 1978*, JO, 6 January 1978 (FDPA) at 2, online: <https://www.cnil.fr/fr/la-loi-informatique-et-libertes#article2>; *EC, Commission Regulation (EC) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*, [2016] OJ, L119/1 (EU GDPR) at 4.

[110] Given that the text of the UK GDPR was essentially derived from the EU GDPR when adopted domestically in the UK, please refer to subsection A of Section III of this report for the application of the EU GDPR in France considering the necessary adaptations.

[111] European Commission, *supra* note 109 at 4.

[112] For instance, articles 5 and 6 of the FDPA essentially replicate articles 6 and 9, respectively, of the EU GDPR. It should be noted, however, that France did not implement the provisions related to transfers of personal data to third countries or international organizations domestically when it comes to research purposes.

[113] French Parliament, *supra* note 109 at 64-77.

[114] French Parliament, *supra* note 109 at 5-6.

[115] Becker, *supra* note 83.

[116] art. L1122-1-1 *Code de la santé publique*.

[117] Only personal data is subject to the FDPA, which excludes viral sequence data with minimal metadata given its unlikelihood of allowing for re-identification.

Provided that the personal health data falls within the scope of the exceptions of Article 9 of the EU GDPR, it must also respect the specific provisions of the FDPA on the processing of personal data concerning health, especially when it comes to research.[118] Under the FDPA, personal health data is subject to two formalities: a declaration attesting in advance that the processing conforms to a standard and, in the case that it does not, an authorization from the *Commission nationale de l'informatique et des libertés* (CNIL).[119] That said, as stated by articles 65 and 67 of the FDPA, some cases are exempt from these formalities. For instance, it is the case of health data processed for internal research purposes.[120] Similarly, no formalities under the FDPA are required for the "processing of personal data concerning health carried out by the organizations or departments entrusted with a public service task contained on a list established by decree of the ministers responsible for health and social security, passed following an opinion of the CNIL, whose sole purpose is to respond, in the event of an emergency, to a health alert and to manage the consequences".[121] Moreover, particular to the context of the COVID-19 pandemic, the French government implemented new laws and decrees to ensure a balance between the need for data protection and the needs associated with research purposes in terms of data access (e.g., sharing data without consent in the context of the pandemic, but restricted to specific organisms authorized by decrees with the CNIL's approval).[122]

B. Concrete actors and practices

The EMERGEN project (*Consortium pour la surveillance et la recherche sur les infections à pathogènes EMERgents via la GENomique microbienne*) is the central consortium for the collection and sharing of SARS-CoV-2 viral sequences in France.[123] The project is coordinated by Public Health France and ANRS-MI, collaboratively tasked to deploy a genomic surveillance system for COVID-19.[124] EMERGEN has three

---

[118] European Commission, *supra* note 109 at 6.

[119] French Parliament, *supra* note 109 at 66; Manon De Fallois, "Les traitements de données de santé à des fins de recherche liés à la COVID-19 : quelle régulation par la CNIL ?" (2021) 29 JDSAM.

[120] "Recherche médicale : quel est le cadre légal ?" (10 December 2018), online: Commission nationale de l'informatique et des libertés (CNIL) <https://www.cnil.fr/fr/recherche-medicale-quel-est-le-cadre-legal>. Research will be deemed internal if it is for the exclusive use of the medical staff.

[121] "Traitement de données à caractère personnel pour la surveillance de l'épidémie de COVID-19" (June 2021), online: Santé Publique France <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/traitement-de-donnees-a-caractere-personnel-pour-la-surveillance-de-l-epidemie-de-covid-19>; The processing of such personal information is subject to the sole provisions of Section 3 of Chapter IV of the EU GDPR.

[122] *Loi n° 2020-546 du 11 mai 2020*, JO, 12 May 2020; *Décret n° 2020-551 du 12 mai 2020 relatif aux systèmes d'information*, JO, 13 May 2020.

[123] "Consortium EMERGEN" (Last modified: 11 January 2022), online: Santé Publique France <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/consortium-emergen>.

[124] *Ibid*.

key goals.[125] First, identify and track the share of variants circulating on the French territory via Flash surveys.[126] Then, identify new variants through the analysis of sequencing results and associated metadata.[127] Finally, EMERGEN looks to promote health research.[128] On this last point, EMERGEN contributes to research on COVID-19 by facilitating access to both non-identifiable data such as viral sequence data with minimal metadata as well as more sensitive data such as host sequence data and viral sequence data with extensive metadata.[129] The consortium gets its data from positive samples (e.g., RT-PCR tests) collected by hospital virology laboratories, medical biology laboratories or general practitioners in the Sentinel network (réseau Sentinelles).[130] The information is then sent to four virology platforms with high-throughput sequencing capabilities (two laboratories of CNR Virus des infections respiratoires (Laboratoires de l'Institut Pasteur & des Hospices Civils de Lyon) and two laboratories of CNR – Laboratoires experts pour l'appui au séquençage du SARS-CoV-2 (le CHU Henri Mondor (AP-HP) & le pôle infectieux de l'APHM à Marseille)).[131] Once sequenced, the data is stored in EMERGEN-DB[132] and harmonized to be sent to international public repositories such as GISAID and ENA.[133] This data is then publicly available in accordance with the data access conditions stated by GISAID and ENA.[134] The metadata usually transmitted to public repositories includes sample number, year of birth of the patient, department of residence, indication of sequencing, date of sample, name and postal code of the sampling laboratory, name of the laboratory in charge of the sequencing, date of receipt of the sampling for sequencing, date of rendering of the result, type of sequencing and its result.[135]

The main data protection authority is the CNIL, which is the French administrative independent regulatory body in charge of ensuring that data privacy laws are duly complied with while personal data is collected,

---

[125] De Fallois, *supra* note 119.

[126] "Enquêtes Flash : évaluation de la circulation des variants du SARS-CoV-2 en France" (March 2022), online: Santé Publique France <https://www.santepubliquefrance.fr/etudes-et-enquetes/enquetes-flash-evaluation-de-la-circulation-des-variants-du-sars-cov-2-en-france>.

[127] *Supra* note 123.

[128] *Supra* note 123.

[129] *Supra* note 123.

[130] Supra note 123.

[131] *Supra* note 123.

[132] EMERGEN-DB is a bioinformatics platform managed by Public Health France and the Institut Français de Bioinformatique (IFB) to host and share all the data from sequencing. The database is accessible to Public Health France's epidemiologists and researchers in a secure digital space.

[133] *Supra* note 123.

[134] *Supra* note 1; "Data availability policy", online: European Nucleotide Archive (ENA) <https://www.ebi.ac.uk/ena/browser/about/policies>.

[135] *Supra* note 121.

stored, and used.[136] This includes supervising and enforcing compliance with the GDPR, the FDPA and other relevant guidelines.[137] Other agencies such as Public Health France, Health Data Hub, Regional Health Agencies *(Agences Régionales de Santé*)[138] and the Caisse nationale d'assurance maladie (CNAM) are also overseeing compliance with domestic and European privacy laws.[139] During the pandemic, the CNIL has published a guide concerning research on COVID-19 to facilitate and accelerate the formalities for researchers.[140] The CNIL manages all applications to access data on COVID-19.[141] If the research using personal health data falls within the scope of a "standard", a simple declaration attesting in advance that the processing conforms to that "standard" is sufficient.[142] In the case that the research does not fall within a standard, but still falls within a reference methodology (MR),[143] a declaration attesting in advance that the processing conforms to that MR is sufficient. For all other cases, the research will only be possible if authorized by the CNIL.[144]

V. COMPARISON

1. Legal similarities

---

[136] French Parliament, *supra* note 109 at 8.

[137] French Parliament, *supra* note 109 at 8.

[138] "Qu'est-ce qu'une agence régionale de santé" (May 2019), online: Agence régionale de Santé <http://www.ars.sante.fr/quest-ce-quune-agence-regionale-de-sante>; "Notifications d'incidents de sécurité aux autorités de régulation : comment s'organiser et à qui s'adresser ?" (18 May 2020), online: CNIL <https://www.cnil.fr/fr/notifications-dincidents-de-securite-aux-autorites-de-regulation-comment-sorganiser-et-qui-sadresser>. They are also informed of security accidents or violations of personal data and must notify the CNIL of several specific elements (e.g., description of the nature of the personal data breach and the approximate number of individuals affected by it).

[139] French Parliament, supra note 109 at 26; "Délibération n° 2020-044 du 20 avril 2020 portant avis sur un projet d'arrêté complétant l'arrêté du 23 mars 2020 prescrivant les mesures d'organisation et de fonctionnement du système de santé nécessaires pour faire face à l'épidémie de covid-19 dans le cadre de l'état d'urgence sanitaire" (20 April 2020), online: CNIL <https://www.cnil.fr/sites/default/files/atoms/files/deliberation_du_20_avril_2020_portant_avis_sur_projet_darrete_relatif_a_lorganisation_du_systeme_de_sante.pdf>.

[140] "Recherches sur le COVID-19 : la CNIL se mobilise" (26 March 2020), online: CNIL <https://www.cnil.fr/fr/recherches-sur-le-covid-19-la-cnil-se-mobilise>.

[141] *Ibid*.

[142] "Référentiel santé RS-001 : Gestion des vigilances sanitaires" (2019), online: CNIL <https://www.cnil.fr/fr/declaration/rs-001-gestion-des-vigilances-sanitaires>.

[143] "Les traitements nécessitant une déclaration de conformité" (2018), online: CNIL <https://www.cnil.fr/fr/traitements-declaration-conformite?field_norme_numerotation_type_value%5B0%5D=6>.

[144] "Demande d'autorisation" (2018), online: CNIL <https://declarations.cnil.fr/declarations/declaration/declarant.display.action?showDraftPopup=true>.

Australia, the US, the UK and France have all adopted privacy laws to protect personal information.[145] While the exact wording used in those laws may change from one country to another, only information that allows for the identification of an individual falls within their scope. For instance, the *Privacy Act 1988* in Australia applies to "personal information" – that is any information or an opinion about an identified individual – while the Privacy Rule in the US applies to "protected health information".[146] In the UK and France, under the UK GDPR and the EU GDPR, respectively, "personal data" means any information relating to an identified or identifiable natural person.[147] The clear consensus across the countries on the type of information that is protected by law simplifies the analysis to determine whether host sequence data with or without metadata qualifies as information about an identifiable individual under each jurisdiction's law. Given that the human genome sequence information collected, used, or disclosed could likely be traced back to the identity of its author, host sequence data is subject to privacy legislation specific to each country. However, the same does not apply to viral sequence data which, if properly de-hosted, is not identifying. The identifiability of the sequence then depends on the nature/amount of metadata attached to it, if any. While viral sequence data with no, or limited, metadata is most likely not be protected by privacy laws, on the contrary, more extensive metadata that allows for the identification of the individual would fall within the definitions of the laws.[148]

The UK GDPR, EU GDPR and the *Privacy Act 1988* in Australia are all broader than the Privacy Rule in the US. The UK GDPR applies to all personal data processed by "controllers" and "processors" in the UK.[149] The EU GDPR applies to a company or entity which processes personal data as part of the activities of one of its branches established in the EU, regardless of where the data is processed. It also applies to a company established outside of the EU and that is offering goods/services – paid or for free – or is monitoring the behaviour of individuals in the EU.[150] The Australian *Privacy Act* is more restrictive than the UK and EU laws but is broader than the Privacy Rule. It applies to government agencies and private sector organisations with an annual turnover of $3 million or more that "collect, use and disclose" personal

---

[145] See Appendix 1.

[146] *Privacy Act 1988*, *supra* note 6; US Department of Health and Human Services, *supra* note 41.

[147] UK Parliament, *supra* note 75; European Commission, *supra* 109 note .

[148] See Appendix 1, "Is viral sequence data with minimal metadata subject to any health privacy laws?"; "Is viral sequence data with extensive metadata subject to any health privacy laws?" and "Is host sequence data, with metadata or viral sequence data subject to any health privacy laws?".

[149] UK Parliament, *supra* note 75, at 4; European Commission, *supra* note 109, at 4.

[150] "Who does the data protection law apply to?", online: European Commission <https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/who-does-data-protection-law-apply_en>.

information.[151] The US Privacy Rule is specific to health information and only applies to the "use or disclosure" by a covered entity, that is, a health plan, a health care clearinghouse, or a health care provider.[152] It only applies to health information that has already been collected. The Privacy Rule states that individually identifiable health information is a subset of health information that is "created or received" by a covered entity.[153] Hence, despite the differences in the scope of the laws under each jurisdiction, all legislations cover at the very least, the collection, use and disclosure of personal health information.[154]

Moreover, the privacy laws of each country allow for the processing of identifiable health information without consent.[155] The UK and France are similar in that respect since they are governed by essentially the GDPR. While always an option, if the processing is necessary for the performance of a task carried out in the public interest and for reasons of public interest in the area of public health, consent will not be required. A similar exception is applicable under the US and Australian laws. In the context of use and disclosure of personal health information, the US Privacy Rule waives the need for consent for public health activities, health oversight activities, research and to counter serious threats to health or safety. In Australia, "permitted health situation" allows certain situations for the collection, use or disclosure of personal health information without consent. That is, the collection for the provision of a health service or for research relevant to public health or public safety, the use or disclosure when necessary for research, or the compilation or analysis of statistics, relevant to public health or public safety, the use or disclosure by an organisation of genetic information when the organisation has obtained the information in the course of providing a health service, and finally, disclosure by an organisation of health information when the organisation provides a health service to the individual and the recipient of the information is a responsible person for the individual.

When comparing countries' privacy legislation, only the US federal legislation does not provide for a distinct category for genomic information.[156] The Privacy Rule applies broadly to protected health information and does not explicitly distinguish genomic information such as genetic sequence data from

---

[151] *Supra* note 6.

[152] US Department of Health and Human Services, *supra* note 41.

[153] *Ibid* at s Notice and Other Individual Rights.

[154] See Appendix 1, "Is host sequence data, with metadata or viral sequence data subject to any health privacy laws?".

[155] See Appendix 1, "What are the exceptions to the restrictions for processing personal information provided in the laws?".

[156] See Appendix 1, "Do health privacy laws provide a distinction between genomic information and personal information more broadly?"; It should be noted, however, that the US Common Rule does provide a distinct category for "identifiable biospecimen".

other types of health information. Thus, it is unlikely that this difference in formulation explains any variances between the countries in terms of deposition of genetic sequence data.

2. Legal Differences

While privacy laws are the main subject of analysis when discussing lawful collection, use and disclosure of information, some legally binding ethical obligations may come into play as well, specifically concerning the requirement of consent. Despite all countries waiving the consent requirement under their respective privacy laws for "research", whether explicitly or for reasons of "public interest", specific laws governing the ethics of research on human beings may require the prior consent of participants when it comes to personal information such as host sequence data or extensive metadata. That is what the National Statement on Ethical Conduct in Human Research (NSECHR) in Australia provides.[157] Similarly, the US Common Rule may require consent for research participation even though the consent requirement has been waived by the applicability of a permitted situation of the Privacy Rule. As for France, FPHC makes consent a condition to carry out research involving humans. The UK, therefore, clearly differentiates itself by not adopting any statutes about research on human beings.[158] This may create an additional barrier for researchers when using host sequence data in their research and might deter the sharing of health information to researchers altogether.

When it comes to the legal formalities required by one country for processing personal health information, France poses the most constraints. Compared to the other countries, France is the only one that requires additional formalities for the processing of identifiable health-related data by both the provider and the receiver of information at the same time. Indeed, the provisions on the processing of health-related data apply to all controllers that process the information, that is both the disclosing entity and the receiving one. These formalities are overseen by the CNIL and add a specific burden to the processing of personal health data, especially when, in the case of research for COVID-19, prior authorization to pursue the research will have to be obtained.

3. Practical similarities

---

[157] *Supra* note 14.

[158] The UK does however provide non legally binding guidelines that apply to human research. See https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/uk-policy-framework-health-and-social-care-research/.

All the countries analyzed possess at least one consortium in charge of collecting viral sequence data either with minimal or extensive metadata. In the US, the NCBI and the N3C are the repositories tasked with the storing and the collection of viral sequence data with minimal and extensive metadata, respectively. Their equivalents are COG-UK in the UK, EMERGEN in France and AusTrakka in Australia. The collection of information is performed through private and public labs in each country, except for the US where the N3C accepts more sensitive data from organizations on a case-by-case basis. In terms of access to non-identifying information such as viral sequence data with limited metadata, the information is open-access and available to the public directly via each platform as well as through the international repositories, with the exception of Austrakka which provides access through GISAID only.

4. Practical differences

Despite the similarities between the countries, there are also notable differences that could contribute to delays in data sharing.[159] Out of the four countries, the US and France present the most significant differences. Indeed, the US is the only country with two different consortia: one that stores less sensitive data such as viral sequence data with minimal metadata (NCBI) and another for more sensitive data such as viral sequence data with extensive metadata (N3C). This difference in structure may contribute to the less efficient data sharing efforts by the US given that both data types are not centralized through the same consortia.[160]

Moreover, France presents more obstacles to personal data sharing than countries like Australia and the UK. The previously mentioned barrier posed by France's legislation translates into a practical hurdle as well. The more rigorous procedure when processing health-related data is less attractive to labs or researchers who are not incentivized as much compared to other countries to share or use personal health information related to COVID-19.[161] While the US, the UK and Australia all require some sort of formality to be filled by the person who requires access to the data, France's policy is stricter.

[159] See Appendix 2A.

[160] The US ranks 3rd compared to the analyzed countries when looking at the latest median deposition days in Appendix 2A.

[161] Although the numbers presented in Appendix 2A are only related to viral sequence data with limited metadata, they are still relevant for host sequence data. France ranks last compared to the other analyzed countries when looking at the latest median deposition days in Appendix 2A.

Finally, with respect to the collection of host sequence data and viral sequence data with extensive metadata, various methods of collection are employed in various countries.[162] Specific to Australia, the country does not have any centralized data collection or sharing effort for host sequence data.[163] For virus sequence data, the amount of metadata shared by individual public labs to AusTrakka is at the discretion of the labs which can create an efficiency problem.[164] In comparison, COG-UK in the UK and EMERGEN in France both collect and store host sequence data and viral sequence data with extensive metadata. In the US, there are also efforts to aggregate viral sequence data with extensive metadata such as the NIH's N3C initiative.[165] However, much of this is still under active implementation, and the details are subject to change.[166]

CONCLUSION

Concerning the processing of viral sequence data with extensive metadata or an individual's sequence data, there are differences between the countries we studied that could create barriers to data sharing. Stricter formalities, like those of France, that apply when processing health-related data and when conducting research may contribute to discouraging data sharing for COVID-19 research. On a more practical side, the involvement of different consortia according to the sensitivity of the data as it is the case in the US, for example, is not optimal and may in also hinder sharing. Moreover, as shown by our comparison, the absence of a centralized platform for any type of data may also contribute to slowing down data sharing. For host sequence data, giving more flexibility to public health labs to share sensitive metadata is a two-edged sword: on one hand, it may encourage data sharing by removing unnecessary restrictions and, on the other, providing public health labs more control over these data might make their sharing more arbitrary also resulting in sub-optimal results.

When it comes to non-personal data such as de-hosted viral sequence data with minimal metadata, our comparison is inconclusive. All the countries analyzed seem to have very similar legal frameworks and structures to support data sharing. Thus, the variation in the timing and degree of completeness of deposit is likely explained by other factors. For one, cultural elements come into play when it comes to data sharing. Since the beginning of the pandemic, the UK, for instance, has been very keen on rapid data sharing to promote the research on COVID-19. This could be explained by a greater familiarity with open science and

---

[162] See Appendix 2B.
[163] *Ibid*.
[164] *Ibid*.
[165] *Supra* Note 64.
[166] *Ibid*.

the benefits it entails for public health in this country. The difference in performance between the four countries is documented in Appendix 2A. Variance in the level of prior preparation of public health authorities, and consortia, to share data can also explain differences in performance. As the numbers in Appendix 2A have shown, it can take some time before the delay from local sequencing to international sharing in a given country reaches an optimal threshold. While all the countries analyzed rank differently on the timing of deposition, assuming a continuing commitment to data sharing, the numbers may continue evolving in a positive direction (shorter delays) within the next few months until they reach a ceiling. This raises a key question, as we are moving into a new stage of the pandemic where restrictions are progressively being lifted, will governments persist in their support for genomic data sharing for public health research?

Bibliography

Legislation

Austl, Commonwealth, Department of Health, *Privacy Policy* (October 2020).

Austl, Commonwealth, *National Health Security Agreement* (July 2011).

Code de la santé publique, 1953.

Décret n° 2020-551 du 12 mai 2020 relatif aux systèmes d'information mentionnés à l'article 11 de la loi n° 2020-546 du 11 mai 2020 prorogeant l'état d'urgence sanitaire et complétant ses dispositions, JO, 13 May 2020.

EC, Commission Regulation (EC) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data , [2016] OJ, L119/1 (EU GDPR).

Health Insurance Portability and Accountability Act, 1996.

Loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, JO, 6 January 1978 (FDPA).

Loi n°2020-546 du 11 mai 2020 prorogeant l'état d'urgence sanitaire et complétant ses dispositions (1), JO, 12 May 2020.

National Statement on Ethical Conduct in Human Research, 1864962755 2007.

Privacy Act 1988 (Cth), 1988/119.

(Retained EU Legislation) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (UK GDPR).

Secondary Material: Articles

Benson, Dennis A et al, "GenBank" (2002) 30 Nucleic Acids Research 1 17–20.

Becker, Regina et al, "COVID-19 Research: Navigating the European General Data Protection Regulation" (2020) 22 Journal of Medical Internet Research 8, e19799. https://doi.org/10.2196/19799.

De Fallois, Manon, "Les traitements de données de santé à des fins de recherche liés à la COVID-19 : quelle régulation par la CNIL ?" (2021) 29 JDSAM.

Harrison, Ewan, *Data Sharing Inside and Outside of the COVID-19 Genomics UK Consortium* (2021).

Song, et al, "Addressing Privacy Concerns in Sharing Viral Sequences and Minimum Contextual Data in a Public Repository during the COVID-19 Pandemic (pre-print)" COVID-19 ELSI special Frontiers in Genetics.

Electronic Sources

"Qu'est-ce qu'une agence régionale de santé" (May 2019), online: Agence régionale de Santé <http://www.ars.sante.fr/quest-ce-quune-agence-regionale-de-sante>.

AusTrakka & Communicable Diseases Genomics Network, *Framework for data sharing and analysis for SARS-CoV-2 in the AusTrakka system* (2020).

CanCOGeN VirusSeq Data Sharing Committee, "CanCOGeN Pan-Canadian Experience In Data Sharing: Lessons Learned and Recommendations for the Road Ahead", Report [forthcoming in 2022].

"CDC's Role in Tracking Variants" (8 September 2021), online: CDC <https://www.cdc.gov/coronavirus/2019-ncov/variants/cdc-role-surveillance.html>.

"SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance (SPHERES)" (11 February 2020), online: Centers for Disease Control and Prevention (CDC) <https://www.cdc.gov/coronavirus/2019-ncov/variants/spheres.html>.

"AusTrakka: Overview", online: CDGN <https://www.cdgn.org.au/austrakka>.

"National genomics reporting", online: Communicable Diseases Genomics Network (CDGN) <https://www.cdgn.org.au/national-genomics-reporting>.

"Our Mission", online: COVID Human Genetic Effort (CHGE) <https://www.covidhge.com/>.

"Participating Centers Regional Coordinators", online: CHGE <https://www.covidhge.com/participating-centers>.

"Sequencing Hub Leaders", online: CHGE <https://www.covidhge.com/sequencing-hubs>.

"Délibération n° 2020-044 du 20 avril 2020 portant avis sur un projet d'arrêté complétant l'arrêté du 23 mars 2020 prescrivant les mesures d'organisation et de fonctionnement du système de santé nécessaires pour faire face à l'épidémie de covid-19 dans le cadre de l'état d'urgence sanitaire" (20 April 2020), online: CNIL <https://www.cnil.fr/sites/default/files/atoms/files/deliberation_du_20_avril_2020_portant_avis_sur_projet_darrete_relatif_a_lorganisation_du_systeme_de_sante.pdf>.

"Demande d'autorisation" (2018), online: CNIL <https://declarations.cnil.fr/declarations/declaration/declarant.display.action?showDraftPopup=true>.

"Les traitements nécessitant une déclaration de conformité" (2018), online: CNIL <https://www.cnil.fr/fr/traitements-declaration-conformite?field_norme_numerotation_type_value%5B0%5D=6>.

"Notifications d'incidents de sécurité aux autorités de régulation : comment s'organiser et à qui s'adresser ?" (18 May 2020), online: CNIL <https://www.cnil.fr/fr/notifications-dincidents-de-securite-aux-autorites-de-regulation-comment-sorganiser-et-qui-sadresser>.

"Recherche médicale : quel est le cadre légal ?" (10 December 2018), online: Commission nationale de l'informatique et des libertés (CNIL) <https://www.cnil.fr/fr/recherche-medicale-quel-est-le-cadre-legal>.

"Recherches sur le COVID-19 : la CNIL se mobilise" (26 March 2020), online: CNIL <https://www.cnil.fr/fr/recherches-sur-le-covid-19-la-cnil-se-mobilise>.

"Référentiel santé RS-001 : Gestion des vigilances sanitaires", (2019), online: CNIL <https://www.cnil.fr/fr/declaration/rs-001-gestion-des-vigilances-sanitaires>.

"A short history of the COVID-19 Genomics UK (COG-UK) Consortium" (17 December 2020), online: COVID-19 Genomics (COG-UK) Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/a-short-history-of-the-covid-19-genomics-uk-cog-uk-consortium/>.

"About Us | COVID-19 Genomics UK Consortium" (14 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/about/about-us/about-us/>.

"CLIMB | COVID-19 Genomics UK Consortium" (2 February 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/priority-areas/data-linkage-analysis/climb/>.

"Original structure | COVID-19 Genomics UK Consortium" (14 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/about/about-us/original-structure/>.

"Public Data & Analysis | COVID-19 Genomics UK Consortium" (12 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/>.

"Reflections on the achievements of COG-UK | COVID-19 Genomics UK Consortium" (22 January 2021), online: COG-UK Consortium | UK-Wide Genomic Sequencing <https://www.cogconsortium.uk/reflections-on-the-achievements-of-cog-uk/>.

COG-UK Steering Group, *COG-UK Publication and Authorship Policy* (2021).

Austl, Commonwealth, Communicable Diseases Intelligence, *COVID-19 Australia Epidemiology Report* by COVID-19 National Incident Room Surveillance Team, Communicable Diseases Intelligence: Technical Supplement (Australian Government Department of Health, 2021).

"Coronavirus disease (COVID-19) epidemiology reports, Australia, 2020–2021" (8 November 2021), online: Department of Health <https://www1.health.gov.au/internet/main/publishing.nsf/Content/novel_coronavirus_2019_ncov_weekly_epidemiology_reports_australia_2020.htm>.

"Data availability policy", online: European Nucleotide Archive (ENA) <https://www.ebi.ac.uk/ena/browser/about/policies>.

"Who does the data protection law apply to?", online: European Commission <https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/who-does-data-protection-law-apply_en>.

"GISAID - Terms of Use" (16 March 2011), online: GISAID <https://www.gisaid.org/registration/terms-of-use/>.

Harrison, *Data Sharing Inside and Outside of the COVID-19 Genomics UK Consortium* (2021).

Office for Civil, *Summary of the HIPAA Privacy Rule* (2013) US Department of Health and Human Services (Last modified: 2021-07-27 T09:13:59-0400).

Office for Civil, "Summary of the Security Rule" (10 September 2020), online: HHS.gov <https://www.hhs.gov/hipaa/for-professionals/security/index.html> (Last modified: 2021-06-28 T08:59:34-0400).

Office for Human Research Protections (OHRP), "2018 Requirements (2018 Common Rule)" (7 March 2017), online: HHS.gov <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html> (Last modified: 2021-08-06T14:38:48-0400).

"About the National COVID Cohort Collaborative" (1 February 2022), online: National Center for Advancing Translational Sciences <https://ncats.nih.gov/n3c/about>.

"GenBank Overview", online: NCBI <https://www.ncbi.nlm.nih.gov/genbank/>.

"How to submit data to GenBank", NCBI online: <https://www.ncbi.nlm.nih.gov/genbank/submit/>.

"NCBI SARS-CoV-2 Resources", s Explore the Data, online: NCBI <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

"COVID-19 response from Australian privacy regulators" (27 March 2020), online: OAIC <https://www.oaic.gov.au/updates/news-and-media/covid-19-response-from-australian-privacy-regulators>.

"Health and medical research", online: OAIC <https://www.oaic.gov.au/privacy/the-privacy-act/health-and-medical-research>.

"National COVID-19 Privacy Principles" (2 September 2021), online: OAIC <https://www.oaic.gov.au/privacy/guidance-and-advice/national-covid-19-privacy-principles>.

"Privacy in your state", online: OAIC <https://www.oaic.gov.au/privacy/privacy-in-your-state>.

"Rights and responsibilities", online: Office of the Australian Information Commissioner (OAIC) <https://www.oaic.gov.au/privacy/the-privacy-act/rights-and-responsibilities>.

"What we do", online: OAIC <https://www.oaic.gov.au/about-us/what-we-do>.

US, Pittman & Levenberg (Global Legal Group), *International Comparative Legal Guides*.

"Consortium EMERGEN" (Last modified: 11 January 2022), online: Santé Publique France <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/consortium-emergen>.

"Enquêtes Flash : évaluation de la circulation des variants du SARS-CoV-2 en France" (March 2022), online: Santé Publique France <https://www.santepubliquefrance.fr/etudes-et-enquetes/enquetes-flash-evaluation-de-la-circulation-des-variants-du-sars-cov-2-en-france>.

"Traitement de données à caractère personnel pour la surveillance de l'épidémie de COVID-19" (June 2021), online: Santé Publique France <https://www.santepubliquefrance.fr/dossiers/coronavirus-covid-19/traitement-de-donnees-a-caractere-personnel-pour-la-surveillance-de-l-epidemie-de-covid-19>.

"International transfers after the UK exit from the EU Implementation Period" (31 January 2022), online: UK ICO <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/international-transfers-after-uk-exit/>.

"Guide to the UK General Data Protection Regulation (UK GDPR)" (29 March 2022), online: UK ICO <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>.'

"Accessing UKHSA Protected Data" (14 March 2022), online: UKHSA <https://www.gov.uk/government/publications/accessing-ukhsa-protected-data/accessing-ukhsa-protected-data>.

APPENDIX 1

Legal Framework for COVID genomic and health data sharing

Questions/Countries	United States	United Kingdom	France	Australia
Is viral sequence data with minimal metadata[167] subject to any health privacy laws[168]?	No, U.S. health privacy legislation only protects individually identifiable health information (personal information) referred to as "protected health information".[169]	No, U.K. privacy laws only protect "personal information".[170]	No, France's privacy laws only protect personal information.[171]	No, Australian privacy laws only protect personal information.[172]

[167] Here, we take it for granted that minimal metadata has been checked and does not allow for re-identification of individuals. E.g., the combination of the geographic location of the host, host age, host age bin, host gender does not constitute personal information.

[168] Only laws that apply throughout the countries, at a national level, were considered.

[169] *HIPAA Administrative Simplification*, 45 CFR at para 160.101- at para 160.550, at para 164.102- at para 164.106 and at para 164.500- at para164.534 (2013) (HIPAA Privacy Rule); *Federal Policy for the Protection of Human Subjects*, 45 CFR at para 46.101- at para 46.124 (2018) (Common Rule).

[170] *Data Protection Act 2018* (UK), c 12; UK Parliament, *supra* note 75.

[171] *Loi n°78-17 du 6 janvier 1978*, JO, 7 janvier 1878 (FDPA); *EC, Commission Regulation (EC) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation)*, [2016] OJ, L 119/1 (EU GDPR).

[172] *Supra* note 6; *supra* note 23; OAIC, *supra* note 17.

Is viral sequence data with <u>extensive</u> metadata subject to any health privacy laws?	Yes, if the extensive metadata likely allows for either a direct or an indirect reidentification of the individual.[173]	Yes, if the extensive metadata likely allows for either a direct or an indirect reidentification of the individual.[174]	Yes, if the extensive metadata likely allows for either a direct or an indirect reidentification of the individual.[175]	Yes, if the extensive metadata likely allows for either a direct or an indirect reidentification of the individual.[176]
Is host sequence data, with metadata or viral sequence data subject to any health privacy laws?	Yes. When data is being "used and disclosed", the protected health information contained in host sequences and potentially identifying metadata is subject to privacy laws.[177]	Yes. When data is being "processed", the individually identifiable information contained in host sequences and potentially identifying metadata is subject to privacy laws.[178]	Yes. When data is being "processed", the individually identifiable information contained in host sequences and potentially identifying metadata is subject to privacy laws.[179] (ref)	Yes. When data is being collected, used, or disclosed, the individually identifiable information contained in host sequences and potentially identifying metadata is subject to privacy laws.[180]

[173] The analysis of whether the metadata allows for reidentification is jurisdiction specific.
[174] *Ibid.*
[175] *Ibid.*
[176] *Ibid.*
[177] US Department of Health and Human Services, *supra* note 41, at para 160.103.
[178] UK Parliament, *supra* note 75 at 4.
[179] European Commission, *supra* note 109 at 4.
[180] *Supra* note 6, s 6, 16A and 16B.

Do health privacy laws provide a distinction between genomic information and personal information more broadly?	No, under the HIPAA Privacy Rule, genetic information is not distinguished from other types of health information.They are all simply treated as personal health information. [181]	Yes, the processing of "data concerning health", including genetic data, is considered as a special category of personal data.[182]	Yes, the processing of 'data concerning health', including genetic data, is considered as a special category of personal data.[183]	Yes, "health information about an individual" including an individual's genetic information falls under the definition of sensitive information.[184]
Which are the government agencies responsible for the promotion/implementation of data privacy legislation?	Various state and federal authorities are involved. [185]	Centrally, the UK Information Commissioner's Office. [186]	Commission nationale de l'informatique et des libertés (CNIL).[187]	The Office of the Australian Information Commissioner (OAIC) is responsible for *Privacy Act* enforcement. Privacy regulators in Australia (federal, state, territory) have created *National COVID-19 Privacy*

[181] US Department of Health and Human Services, *supra* note 41, at para 160.103 and at para 164.501.
[182] UK Parliament, *supra* note 75, at art.9.
[183] European Commission, *supra* note 109, at 32.
[184] *Supra* note 6.
[185] US Department of Health and Human Services, *supra* note 41.
[186] "About the ICO" (29 November 2021), online: UK ICO <https://ico.org.uk/about-the-ico/>.
[187] French Parliament, *supra* note 109 at 8.

				Principles for a consistent national approach.[188]
What are the exceptions to the restrictions for processing[189] personal information provided in the laws?	a) For public health activities [190] b) For health oversight activities[191] c) For research [192] d) To counter serious threats to health or safety.[193]	a) For reasons of substantial public interest[194] b) For the purposes of preventive or occupational medicine, […], medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems	a) For reasons of substantial public interest[197] b) For the purposes of preventive or occupational medicine, […], medical diagnosis, the provision of health or social care or treatment or the	Information that is collected for COVID-19 related reasons should generally not be used for other purposes.[201] Exceptions exist if use is reasonably necessary and there are protections from inappropriate disclosure: a) Collection—provision of a

[188] *Supra* note 15; OAIC, supra note 17.

[189] Processing should be understood as "usage and disclosure" under U.S. law and "collection, use, or disclosure" under Australian law.

[190] US Department of Health and Human Services, *supra* note 41 at s Public Health Activities.

[191] *Ibid* at s Health Oversight Activities.

[192] *Ibid* at s Permitted Uses and Disclosures.

[193] *Ibid*.

[194] UK Parliament, *supra* note 75 at 6 and 9.

[197] European Commission, *supra* note 109 at 6 and 9.

[201] OAIC, *supra* note 17.

		and services. [195] c) For the reasons of public interest in the area of public health.[196] These exceptions apply if one of the situations of Article 6 of the UK GDPR applies as well.	management of health or social care systems and services.[198] c) For the reasons of public interest in the area of public health.[199] These exceptions apply if one of the situations of Article 6 of the GDPR applies as well. [200]	health service b) Collection—research etc. c) Use or disclosure—research etc. d) Use or disclosure—genetic information e) Disclosure—*responsible person* (e.g., legal guardian, emergency contact) for an individual.[202]

[195] *Ibid*.

[196] *Ibid*.

[198] European Commission, *supra* note 109 at 6 and 9.

[199] European Commission, *supra* note 109 at 6 and 9.

[200] European Commission, *supra* note 109 at 5-6.

[202] Even though consent is not required for the research collection, use or disclosure of sensitive information, consent may still be required to participate in the research study, according to the *National Statement on Ethical Conduct in Human Research* (2007), s 2.3.10; *supra* note 14.

What security restrictions is host data with metadata or viral sequence data subject to?	HIPAA Security Rule implements various security requirements by using appropriate administrative, physical, and technical safeguards to ensure the confidentiality, integrity and security of this information.[203]	Data Controllers and Data Processors have to implement technical and organizational measures that ensure a level of data security appropriate for the level of risk presented by processing personal data.[204]	Data Controllers and Data Processors have to implement technical and organizational measures that ensure a level of data security appropriate for the level of risk presented by processing personal data.[205]	The *Privacy Act* requires that entities that hold personal information and are governed by the *Act* take reasonable steps in the applicable circumstances to protect the personal information from misuse, loss, unauthorized access, disclosure, modification, and use. Personal information should be stored in Australia.[206]

[203] *HIPAA Administrative Simplification*, 45 CFR at para 160.101- at para 160.550, at para 164.102- at para 164.106 and at para 164.302- at para 164.318 (2013) (HIPAA Security Rule).

[204] UK Parliament, *supra* note 75, at 32.

[205] European Commission, *supra* note 109, at 32.

[206] OAIC, *supra* note 17.

APPENDIX 2A

Data flow for viral data with minimal metadata

Questions/Countries	United States	United Kingdom	France	Australia
Median days to deposition (GISAID) since January 10, 2020 [207]	25	10	24	21
Median days to deposition (GISAID) since July 31, 2021[208]	21	9	22	14
Is there a central point of access to data?	Yes	Yes	Yes	Yes[209]
Is there a national and centralized health care system?	No	Yes	Yes	No

[207] Kalia Kishan et al, "The lag in SARS-CoV-2 genome submissions to GISAID" (2021) 39 Nat Biotechnol 9 at 1058–1060.

[208] *Ibid*.

[209] Although Austrakka is the central point of access, the data is not accessible without approval from individual labs. The data however, remains accessible via GISAID.

Which consortia are Centrally involved in the collection of viral sequence data?[210]	**SPHERES** collaborates with local, state, federal, academic, non-profit, and private organizations to create various scientific data standards, but does not collect data. **The NCBI** is an Open-Access data repository based in the U.S. The NCBI Genbank hosts a variety of pathogen viral data types including SARS-CoV-2 sequences.[211]	**COG-UK** is actively collaborating with the 4 public health agencies of the UK (England, Scotland, Wales, Northern Ireland), academic institutions, National Health Service laboratories, and diagnostic laboratories.[212]	**EMER-GEN** labs collect the viral data from public health labs which send it for sequencing the virus may be sequenced from the biological samples of positive cases to detect possible mutations and then transmit them to Public Health France and the INSERM/ANRS-MIE.[213]	**AusTrakka** specifically for COVID-19 viral data, under the Communicable Diseases Genomics Network (CDGN).[214] **The National Incident Room** (NIR) receives monthly COVID-19 reports from AusTrakka.[215]

[210] We focused on country specific consortia rather than international consortia such as GISAID, ENA, ELIXIR Hub.

[211] NCBI, *supra* note 58.

[212] In close collaboration with COG-UK, viral genomic data is stored in the Cloud Infrastructure for Microbial Bioinformatics (CLIMB).

[213] Data is stored on EMERGEN-DB.

[214] June 2021 - data was projected to have been moved from lab-specific servers to being stored in national Azure cloud; AusTrakka & Communicable Diseases Genomics Network, *supra* note 28.

[215] *Supra* note 26.

How is the viral sequence data collected?	Individual labs and authorities deposit data into NCBI.[216]	Public health labs and their affiliates send the data to COG-UK.[217] COG-UK is also involved in depositing such data into international/ continental databases such as GISAID and ENA.[218]	The data is collected directly by labs and transmitted to EMER-GEN labs for sequencing. [219]	AusTrakka: All Australian (and New Zealand) Public Health Labs have access to AusTrakka and upload sequences.[220]
Are there any restrictions on data access?	No, viral sequence data is public and open access.[221] Data access is available through several resources,	No, viral sequence data is public and open access. [223]	No, viral sequence data is public and open access.[224]	Viral sequence data is public and open access through GISAID, but is not publicly available through AusTrakka.[225]

[216] Individual labs will also deposit data into the GISAID. The CDC is also sometimes involved (E.g., data curation, deposition) in data deposition to international data repositories such as GISAID, and the NCBI when labs lack such capacity.

[217] Data is stored in CLIMB.

[218] COG-UK: https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/; GISAID: https://www.epicov.org/; ENA: https://www.ebi.ac.uk/ena/browser/view/PRJEB37886?show=reads.

[219] *Supra* note 123.

[220] Sequences uploaded to GISAID by Public Health Laboratories are automatically captured by AusTrakka; *supra* note 29.

[221] Through NCBI and GISAID.

[223] *Supra* note 86; *supra* note 89.

[224] *Supra* note 1; "Data availability policy", online: European Nucleotide Archive (ENA) <https://www.ebi.ac.uk/ena/browser/about/policies>.

[225] Through GISAID only.

	GISAID and the NCBI being most prominent. [222]			

[222] *Supra* note 1; *supra* note 58.

APPENDIX 2B

Data flow for viral data with extensive metadata or host sequence data

Questions/Countries	United States	United Kingdom	France	Australia
Is there a central point of access?	Yes	Yes	Yes	No[226]
Is there a national and centralized health care system?	No	Yes	Yes	No
Which consortia are involved in the collection of data?	**The National COVID Cohort Collaborative (N3C) project** is a part of the larger NIH National Center for Advancing Translational Sciences	**COG-UK** is the central consortium involved in the collection of data. COG UK also collaborates with the 4 public health	**EMER-GEN** is the central consortium responsible for viral genomic surveillance in France.[230] EMER-GEN is coordinated by Public Health France ANRS/Maladies	**AusTrakka** specifically for COVID-19 viral data with metadata, under the Communicable Diseases Genomics Network

[226] There is no centralized process for collecting host sequence data. Individual public health labs may collect viral data with extensive metadata or host sequence data.

[230] According to https://www.santepubliquefrance.fr, the consortium has a bioinformatic platform available to host and share all sequencing data, EMERGEN-DB, developed and maintained by the Institut Français de Bioinformatique. It allows the analysis and availability of data from several thousand genomes produced each week. This database is accessible to Santé publique France epidemiologists and researchers in a secure digital space.

	(NCATS) and stores viral sequence data with extensive metadata. As of January 2022, such data is not available, but efforts are proceeding. Furthermore, this effort will only include the records of a fraction of individuals.[227]	agencies of the UK (England, Scotland, Wales, Northern Ireland), academic institutions, National Health Service laboratories, and diagnostic laboratories to rapidly deposit sequencing and the associated metadata in controlled-access repositories.[228] For external (non-UK) researchers, there is controlled access via the UKHSA. [229]	Infectieuses émergentes. Data is deposited by EMER-GEN in GISAID and the ENA.[231]	(CDGN).[232]
How is the viral sequence data collected?	The N3C accepts user data submission on a	Public health labs and their affiliates send the data to	It is collected by hospital virology laboratories, medical	Individual public health labs collect data and

[227] *Supra* note 64.

[228] *Supra* note 88. In close collaboration with COG-UK, viral genomic data is stored in the Cloud Infrastructure for Microbial Bioinformatics (CLIMB).

[229] *Supra* note 101.

[231] *Supra* note 123.

[232] June 2021 - data was projected to have been moved from lab-specific server to being stored in the national Azure cloud; AusTrakka & Communicable Diseases Genomics Network, *supra* note 28.

	case-by-case basis. The N3C team must assess that the submitted data meets legal and regulatory requirements. Several primary data collecting organizations are collaborating by making their data available to the N3C effort.[233]	COG-UK. UKHSA collaborates with the DHSC, academic partners, industry, and other governmental organizations to collect data.[234]	biology laboratories or general practitioners of the "réseau Sentinelles". As a part of the public health effort, positive biological samples are collected based on pathogen surveillance.[235]	submit viral data with metadata to AusTrakka – the amount of metadata submitted is the decision of the individual public health lab.[236]
Are there any restrictions on data access?	Yes, depending on the sensitivity of the data, it would fall either under "synthetic", "de-identified" or "limited",	Yes, depending on the sensitivity of the data, it would fall either under "consortium" or "restricted", which both provide for higher	Yes, public health data may be used for research purposes if authorized by the CNIL. Research projects must apply for data access.[239]	Yes, accessing viral sequence data/metadata requires the permission of the individual public health lab that maintains

[233] *Supra* note 64.

[234] *Supra* note 101.

[235] *Supra* note 123.

[236] AusTrakka & Communicable Diseases Genomics Network *supra* note 28.

[239] The CNIL sets different baseline methodologies, such as MR-005 and MR-006, access in the context of public interest and necessary to pursue a legitimate interest, respectively, see "Référentiel santé RS-001 : Gestion des vigilances sanitaires" (2019), online: CNIL <https://www.cnil.fr/fr/declaration/rs-001-gestion-des-vigilances-sanitaires>; *supra* note 140; *supra* note 143.

	which provide for higher restrictions in terms of access.[237]	restrictions. Yes, the UKHSA requires researchers to submit a data access request, and then imposes data access requirements via a legally binding data sharing contract.[238]		custodianship of the data.[240]
What are the measures taken to protect the data?	All work must be done within the enclave, and no data may be downloaded.[241] NCATS has established a COVID-19 Data Transfer	All this data is combined with the viral sequences in COG-UK's central database, called CLIMB-COVID-19.[243] The data in this database is de-personalized.	Within the framework of SARS-CoV-2 genomic surveillance, positive samples (RT-PCR) to be sent to the various sequencing laboratories must be	Viral sequence with extensive metadata is not accessible to the public via AusTrakka. Requests must be approved by the individual public health

[237] *Supra* note 70.

[238] Supra note 101.

[240] AusTrakka & Communicable Diseases Genomics Network, *supra* note 28.

[241] The N3C Data Enclave's secure, cloud-based environment is certified through the Federal Risk and Authorization Management Program, or FedRAMP, which provides standardized assessment, authorization and continuous monitoring for cloud products and services, ensuring the validity of the data while protecting patient privacy.

[243] CLIMB-COVID is a secure digital cloud infrastructure that integrates (alignment with consensus sequences) collected viral sequences and extensive metadata. Data analysis and storage in this secure infrastructure, which prevents unauthorized sharing, data tampering, etc.

	Agreement (DTA) that provides terms and conditions for data transfer and outlines the general terms of data use. N3C data may be used only for COVID-19 research purposes. Before researchers can request access to the data, their institutions must execute a Data Use Agreement (DUA) with NCATS.[242]	The CLIMB-COVID-19 system meets the security standards for processing NHS data.[244] the UKSHA data sharing agreement will specify the measures that must be taken on a case-by-case basis.[245]	accompanied by a shipping note.[246] For internal research projects, no formalities are required. The data controller must include the relevant data processing in its register of processing activities.[247] If the research project falls into a reference methodology[248] as set out by the CNIL, a declaration of compliance with the	labs that have custodianship.[250]

[242] *Supra* note 64.

[244] See here for more details on the mentioned NHS requirements: https://www.dsptoolkit.nhs.uk/.

[245] *Supra* note 101.

[246] *Supra* note 123.

[247] For more information, *supra* note 120.

[248] A reference methodology therefore ensures that the rights and freedoms of the persons concerned are respected, while promoting research by providing legal certainty to researchers. It is a legal and technical "frame of reference" specific to research, studies, and evaluations in the health field. It allows organizations that comply with it to carry out their study without first seeking authorization from the CNIL, thanks to a compliance undertaking. See "Santé : la CNIL organise une consultation publique sur la modification des méthodologies de référence pour la recherche en santé" (10 January 2022), online: CNIL <https://www.cnil.fr/fr/sante-la-cnil-organise-une-consultation-publique-sur-la-modification-des-methodologies-de-reference>.

[250] AusTrakka & Communicable Diseases Genomics Network, *supra* note 28.

			corresponding reference methodology must be completed. If the research project does not fall into a reference methodology as set out by the CNIL, an application for "research" authorization must be submitted in addition to the elements usually required.[249]	

[249] Such as opinion of the competent committee, research protocol and its summary, information document intended for patients, etc; *supra* note 120.