Interoperable medical data: The missing link for understanding COVID-19

Denis C Bauer, Alejandro Metke-Jimenez, Sebastian Maurer-Stroh, Suma Tiruvayipati, Laurence Wilson, Yatish Jain, Amandine Perrin, Kate Ebrill, David P Hansen, Seshadri S Vasan

Received: 11 July 2020 | Revised: 14 October 2020 | Accepted: 20 October 2020

DOI: 10.1111/tbed.13892

RAPID COMMUNICATION

Transboundary and Emerging Diseases WILEY

Interoperable medical data: The missing link for understanding COVID-19

Denis C. Bauer[1,2] | Alejandro Metke-Jimenez[3] | Sebastian Maurer-Stroh[4,5,6,7] | Suma Tiruvayipati[7,8,9] | Laurence O. W. Wilson[1] | Yatish Jain[1] | Amandine Perrin[10,11,12] | Kate Ebrill[3] | David P. Hansen[3] | Seshadri S. Vasan[13,14]

[1]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Geelong, Australia, Australia

[2]Department of Biomedical Sciences, Macquarie University, Macquarie Park, NSW, Australia

[3]Commonwealth Scientific and Industrial Research Organisation, Australian e-Health Research Centre, Herston, QLD, Australia

[4]Agency for Science Technology and Research, Bioinformatics Institute, Singapore, Singapore

[5]Department of Biological Sciences, National University of Singapore, Singapore, Singapore

[6]National Public Health Laboratory, National Centre for Infectious Diseases, Ministry of Health, Singapore, Singapore

[7]Global Initiative on Sharing All Influenza Data (GISAID), Munich, Germany

[8]Infectious Diseases Programme, Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[9]Bacterial Genomics Laboratory, Genome Institute of Singapore, Singapore, Singapore

[10]Bioinformatics and Biostatistics Hub, Department of Computational Biology, Institut Pasteur, USR 3756 CNRS, Paris, France

[11]Microbial Evolutionary Genomics, Institut Pasteur, UMR 3525 CNRS, Paris, France

[12]Collège doctoral, Sorbonne Université, Paris, France

[13]Australian Centre for Disease Preparedness, Commonwealth Scientific and Industrial Research Organisation, Geelong, VIC, Australia

[14]Department of Health Sciences, University of York, York, UK

Correspondence
Professor Seshadri S. Vasan, CSIRO Australian Centre for Disease Preparedness, 5 Portarlington Road, Geelong, VIC 3220, Australia
Email: vasan.vasan@csiro.au

Funding information
Genome Institute of Singapore (S.T.); Institut Pasteur (A.P.); Temasek Foundation (S.T.); Agency for Science, Technology and Research (S.T.; S.M-S.); Coalition for Epidemic Preparedness Innovations (S.S.V.)

Abstract

Being able to link clinical outcomes to SARS-CoV-2 virus strains is a critical component of understanding COVID-19. Here, we discuss how current processes hamper sustainable data collection to enable meaningful analysis and insights. Following the 'Fast Healthcare Interoperable Resource' (FHIR) implementation guide, we introduce an ontology-based standard questionnaire to overcome these shortcomings and describe patient 'journeys' in coordination with the World Health Organization's recommendations. We identify steps in the clinical health data acquisition cycle and workflows that likely have the biggest impact in the data-driven understanding of this virus. Specifically, we recommend detailed symptoms and medical history using the FHIR standards. We have taken the first steps towards this by making patient status mandatory in GISAID ('Global Initiative on Sharing All Influenza Data'), immediately resulting in a measurable increase in the fraction of cases with useful patient information. The main remaining limitation is the lack of controlled vocabulary or a medical ontology.

Denis C. Bauer and Alejandro Metke-Jimenez: Co-first authors.

David P. Hansen and Seshadri S. Vasan: Co-corresponding authors.

KEYWORDS
COVID-19, genome sequence, GISAID, ontology, patient information, SARS-CoV-2

1 | INTRODUCTION

Being able to link clinical outcomes to virus strains is a critical component of understanding COVID-19; however, current data collection practices hamper such analyses and require updating to support robust insights gained from the data collected.

GISAID, established originally as the Global Initiative on Sharing All Influenza Data (Elbe & Buckland-Merrett, 2017), has widened its remit with the EpiCoV™ database to become the principal platform for the sharing of genomic sequences of SARS-CoV-2 (hCoV-19) from around the world. Such convergence by the global scientific community around a single database is critical to permit a near-real-time analysis of how the virus is evolving. While currently only 1 out of 258 confirmed cases (Worldometers Coronavirus, n.d.) sees the virus sequence submitted (i.e. 36,080,088 COVID-19 cases and 139,967 published SARS-CoV-2 sequences as of 1 October 2020, which indicates that circa 1 out of 258 cases are sent for virus sequencing), it represents the most thorough surveillance of an emerging virus outbreak in history (Massive coronavirus sequencing efforts urgently need patient data - Nature India, 2020).

It is therefore critical to supplement the collected information on the virus genomes with the other critical component informing patient outcome: medical information. Such de-identified patient data would provide the missing information that enables the virus evolution to be linked to its host's clinical factors. For example, several studies have suggested the emergence of virus isolates associated with greater in vitro titres and cytopathic effects (Yao et al., 2020); greater infectivity (Korber et al., 2020); greater transmissibility (McAuley et al., 2020); and similar (Zhang et al., 2020) or attenuated (Su et al., 2020) phenotypes with consequent outcomes.

Such observed variations, especially disease severity and phenotypic changes, may be attributable to genomic evolution and adaptation to the new human host. However, current analyses are confounded by factors such as co-morbidities, capacity of the healthcare system in terms of diagnostic testing, treatment choices and reporting of severity and fatality—making it impossible to robustly link patient outcome to genomic changes in the virus. This limits studies to being merely observational by reporting genomic differences of the virus (Bauer et al., 2020) or inferring pathogenicity from cell culture measurements such as replication rate (Yao et al., 2020) and cell toxicity (Chu et al., 2020). While such in silico and in vitro studies are insightful, they are not a reliable predictor of disease severity in vivo.

Recognizing the need for clinical data, GISAID enables 'patient status' to be recorded for each submitted isolate and made this field mandatory as of 27 April 2020. Two snapshots were taken to assess the uptake of this feature. One month after the change (15 May 2020), only 3% provided relevant information for this field, for instance, 9% (506/5122) of submitted isolates have this field filled in and of these only a third (164) have provided clinical information (Figure 1a). At the 6-month mark (01 October 2020), this increased to 13% of entries with data other than 'unknown' (15,907/125,654); however, the usefulness of this data remains variable (Figure 1b). The word clouds highlight that 'unknown' remains the largest fraction and that the free-text field gives rise to a wide range of different descriptions identifying the same status.

There are hence two areas where current processes hamper sustainable and meaningful data collection. Firstly, information is currently not captured in a standardized form that is tailored to COVID-19 infections; secondly, patient information is frequently not available when genomic information is submitted, and workflows are not set up to amend entries retrospectively.

2 | CAPTURING CLINICAL DATA IN STANDARDIZED FORMS

Data that are collected and submitted to a central repository such as GISAID likely come from multiple sources, with consequently a wide range of digital-readiness levels. For example, it might be extracted from Electronic Medical Records (EMRs) where the data are already in a structured form. However, it may also be that relevant information needs to first be extracted out of digital- or paper-based clinical notes. In the latter case, the same clinical symptom might be described differently, complicating downstream reporting or grouping of records. Hence, converting clinical observations into standardized terms, so called clinical terminologies that are applicable across the world, is relevant. Figure 2 illustrates this problem on the concept 'loss of sense of smell', which has several synonyms, such as 'anosmia' and 'absent smell', but is represented as a single concept in the 'SNOMED CT' (Systematized Nomenclature of Medicine Clinical Terms) terminology.

While the progression towards EMRs is a much larger, multilayer problem that cannot be addressed quickly amid a pandemic, the mode of primary data collection into the central repository can be controlled by introducing standardized fields implementing standardized terminologies. This would ensure that researchers have a computable set of data to build robust statistical methodologies and artificial intelligence-based analyses, gaining insights from genomic and clinical data.

However, there are several clinical terminologies, such as Systematized Nomenclature of Medicine (SNOMED) and International Classification of Diseases (ICD). SNOMED CT is the most comprehensive multilingual health terminology in the world,

FIGURE 1 Word cloud of GISAID 'patient status' entries, where word size represents number of entries with this term (log10-transformed and pseudocounts to also visualize low frequency). (a) snapshot from 15 May 2020, (b) snapshot from 1 October 2020, after 'unknown' was made the default status when no status is provided. Actual counts are in Table S1; typographical and other errors faithfully reproduced, though now corrected in GISAID

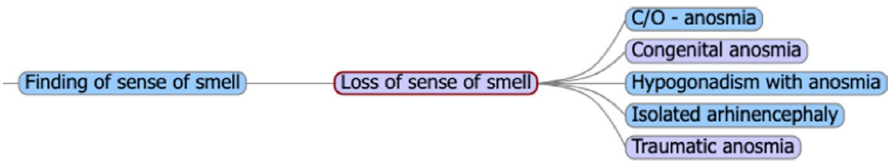

FIGURE 2 Example of a hierarchical terminology relationship

while ICD is a classification specializing on disease description. The main difference between them is that SNOMED CT is much more detailed and can be used to capture fine-grained clinical information while ICD is primarily a classification designed for reporting.

In addition to clinical terminologies, a standard that defines which clinical data should be collected is also needed. For example, in this case it is useful to capture symptoms, risk factors and complications, among others. This is usually referred to as the *information model*. The new 'Health Level Seven' (HL7) standard called 'Fast Healthcare Interoperable Resource' (FHIR) stands out as the best choice, given its substantial uptake and excellent support for clinical terminologies.

2.1 | Emerging standardization for COVID-19

There are multiple efforts that currently aim to define the minimal COVID-19-relevant clinical data.

The World Health Organization (WHO) has developed a case-based reporting form and data dictionary, as well as interim guidance to clinicians regarding case definitions and clinical syndromes associated with COVID-19 (Table 1). Although the WHO's forms are more likely to be accepted by clinical teams around the world, the resulting forms do not capture clinical symptoms and outcomes in detail, for example, only a field for indicating if the patient was showing symptoms but not which symptoms. Similarly, clinical course and outcomes are captured in little detail.

Patient State	Descriptor	Score
Uninfected	Uninfected; no viral RNA detected	0
Ambulatory mild disease	Asymptomatic; viral RNA detected	1
	Symptomatic; independent	2
	Symptomatic; assistance needed	3
Hospitalised: moderate disease	Hospitalised; no oxygen therapy*	4
	Hospitalised; oxygen by mask or nasal prongs	5
Hospitalised: severe diseases	Hospitalised; oxygen by NIV or high flow	6
	Intubation and mechanical ventilation, pO_2/FiO_2 ≥150 or SpO_2/FiO_2 ≥200	7
	Mechanical ventilation pO_2/FIO_2 <150 (SpO_2/FiO_2 <200) or vasopressors	8
	Mechanical ventilation pO_2/FiO_2 <150 and vasopressors, dialysis, or ECMO	9
Dead	Dead	10

FIGURE 3 Minimal common outcome measure as compiled by WHO. Figure reproduced from WHO Working Group on the Clinical Characterisation & Management of COVID-19 infection, 2020

TABLE 1 Web resources for the standardized capture of COVID-19 information

Initiative	Target audience	Description	Link
WHO	Clinicians and health authorities	COVID−19 case-based reporting form, data dictionary, case definitions and clinical syndromes	https://apps.who.int/iris/rest/bitstreams/1270897/retrieve, https://www.who.int/docs/default-source/coronaviruse/2020-02-27-data-dictionary-en.xlsx?sfvrsn=9dbd9418_6&download=true, https://www.who.int/publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected
COVID−19 host genetics initiative	General public	Questionnaire capturing symptoms and co-morbidities	https://docs.google.com/spreadsheets/d/1RXrJIzHKkyB8qx5tHLQjcBioiDAOrQ3odAuqMS3pUUI/edit#gid=0, https://docs.google.com/document/d/1eMdzhO5xk-MACxjz-kOUJLP6Jort5KuwoOa_u-aZPHs/edit
COVID−19 host genetics initiative	Pathology/clinical data curators	Relevant IC10D and SNOMED terms	https://drive.google.com/file/d/1ck0ABYZ6oYnMStoYnGpnA7n1W6wcY3_6/view
SNOMED	Developers	COVID−19 vocabulary	http://snomed.org/cv19
ICD10	Developers	COVID−19 vocabulary	https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf?ua=1
FHIR	Developers	COVID−19 vocabulary	https://docs.google.com/spreadsheets/d/1P3DgnLOvr31H4cIfRa_cTfkdhBCC8acTzPCbHmjakrI/edit?usp=sharing, https://covid-19-ig.logicahealth.org/index.html
CSIRO	Pathology/clinical data curators	Implementation Guide for genomic and patient data collection	https://genomics.ontoserver.csiro.au/covid19/UserInterfaceConsiderations.html

Aiming to capture more details and interpret their clinical impact, the WHO has compiled a common outcome measure that groups patients into 5 categories ('Uninfected', 'Ambulatory mild disease', 'Hospitalized modest disease', 'Hospitalized severe disease' and 'Dead', as illustrated in Figure 3) using a range of clinical data (WHO Working Group on the Clinical Characterisation & Management of COVID-19 infection, 2020).

However, achieving international agreement on the exact thresholds for the grouping is likely difficult, especially as new evidence about the severity of individual symptoms becomes available (Menni et al., 2020). It might hence be a more prudent approach to capture symptoms directly, as taken by the COVID-19 host genetics initiative (The COVID-2020 Host Genetics

Initiative, 2020), which aims to annotate existing human genomic information in large BioBanks by collecting self-reported COVID-19 status from its participants. This consortium has put together a questionnaire aimed at capturing COVID-19 symptoms and co-morbidities, which may provide a way to capture the disease status directly from the patient.

Worldwide standards for classifications and terminologies have been updating the content to include concepts and terms that describe or classify COVID-19-related diseases and symptoms. A clinical diagnostic dictionary looking at the collection of these terms was put together for the COVID-19 host genetics initiative, collecting terms from both ICD10 and SNOMED (see Table 1).

This highlights the different approaches the two vocabularies have taken. ICD10 opted for a high-level 'COVID-19' term to enable counting of the number of COVID-19 cases, while SNOMED International is adding several COVID-19-related diagnosis codes to SNOMED CT, providing the ability to capture more specific data about the impact of the disease. Note that SNOMED CT allows for these cases to be grouped and cases counted.

There are also initiatives to develop data models for sharing COVID-19 clinical data using the 'Fast Healthcare Interoperable Resource' (FHIR) standard from HL7 International. One such example is from Logical Health, a consortium of healthcare providers and technical companies in the USA. The FHIR Implementation Guide provided by Logical Health is a resource for capturing information to help with the treatment of patients in hospital.

2.2 | What could interoperability look like for COVID-19

Using existing technology and incorporating the above discussed guidelines for COVID-19 symptoms and severity, we built an example FHIR Implementation Guide (FHIR IG) and implemented it as a FHIR questionnaire (see Table 1). This allows the flexible collection of relevant terms for a specific use case and allows them to be expressed as an input form for data collection, for example into GISAID. Unlike the FHIR IG from Logica, which focuses on patient care, patient screening, public health reporting and general research, we designed the questionnaire (fields and values) for the specific use case of linking genomic data with clinical outcomes.

The FHIR IG captures the following types of information:

- Demographic information—such as the age and gender of the patient
- Basic clinical information—such as blood type
- Pre-existing clinical information—such as co-morbidities and medication
- Travel history
- Observed COVID Symptoms
- Severity of COVID disease
- Outcome
- Immunization history

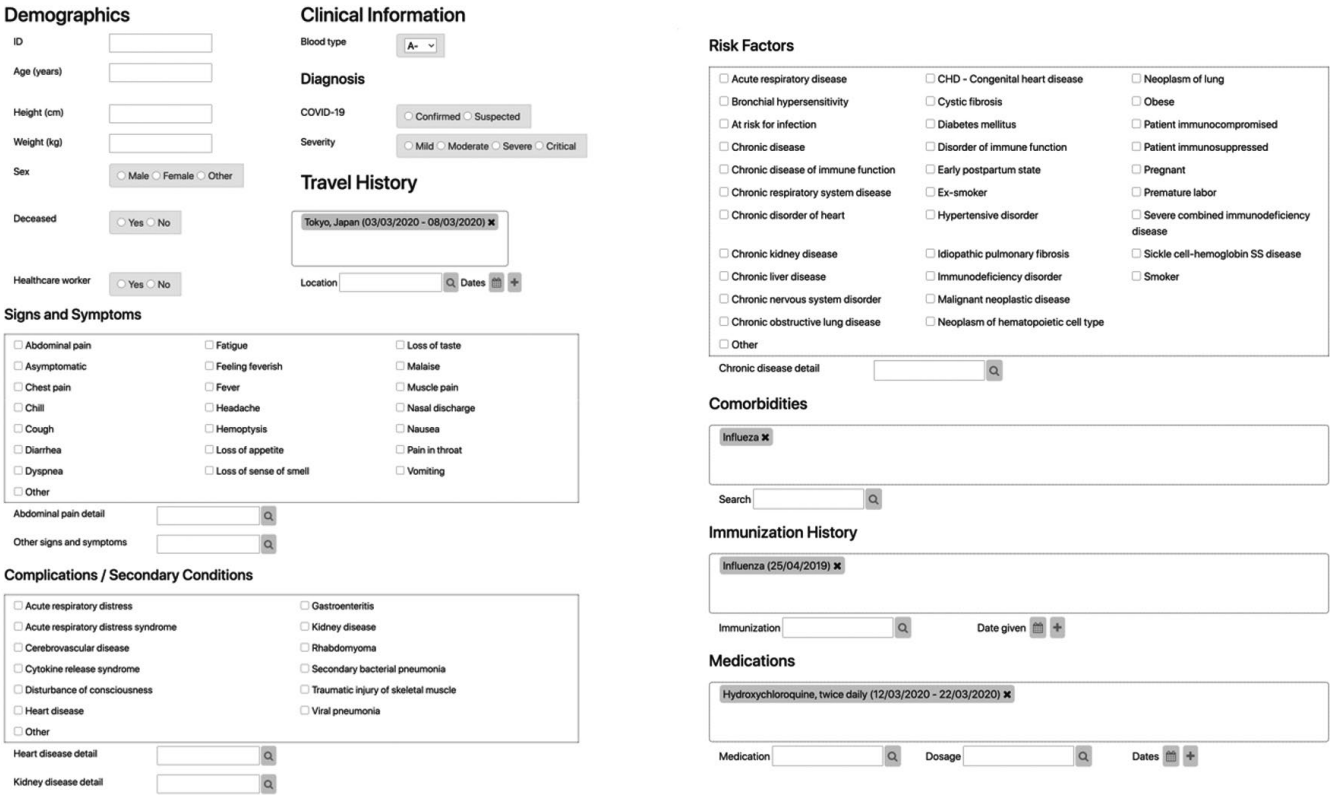

FIGURE 4 Example entry form for COVID-19 patient information given in the Implementation Guide

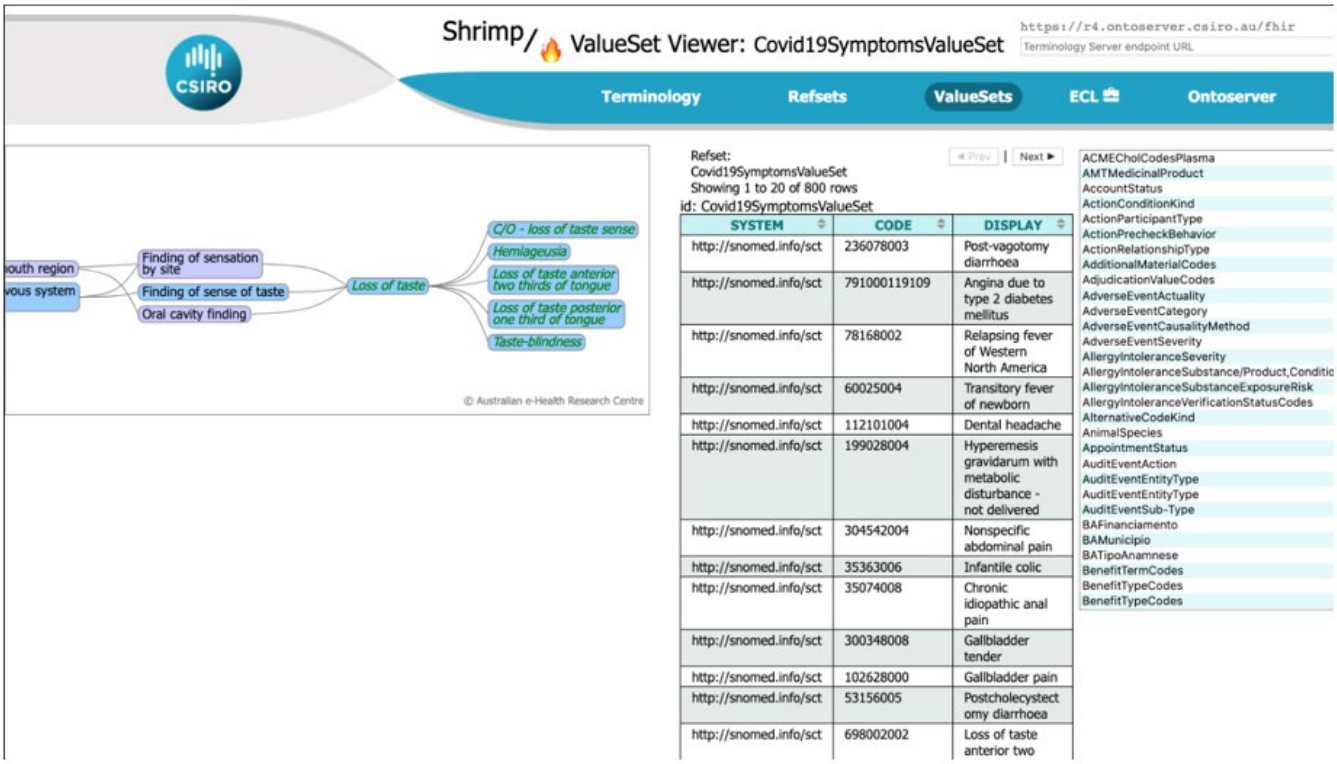

FIGURE 5 SNOMED CT COVID-19 symptoms value set shown in the Shrimp browser

The FHIR IG provides a set of standard terms from the SNOMED CT clinical terminology in the form of value sets. These are available in the documentation as well as programmatically from a clinical terminology service. Advice around the design of a user interface is also provided—with an example of an implementation for the form used to collect the information shown in Figure 4D.

The FHIR IG provides the guidance needed to build different approaches to data collection. For example, one approach might be to use data extracted from an Electronic Medical Record (EMR) system or a research Electronic Data Capture (EDC) system like REDCap (Harris et al., 2019) for sharing with an organisation such as GISAID. There are existing tools that can be used to facilitate this transformation (Metke-Jimenez & Hansen, 2019). Alternatively, a specific cloud-based web form can be built to capture data and store it in a cloud-based FHIR repository for later analyses.

The value sets developed for the different fields in the clinical entry form can be browsed using a terminology browser. Figure 5 shows the symptom-value set in the CSIRO Shrimp browser, a front end for CSIRO's terminology server Ontoserver (Metke-Jimenez et al., 2018).

3 | CLINICAL WORKFLOWS NEED TO REVISIT ENTRIES

While GISAID enables updates to submitted entries as more patient data become available, updating a submitted entry with clinical information is currently not a wide-spread practice. This in part is due to privacy restriction having prevented the sharing of patient information (Dyer, 2020). While the current content of GISAID was carefully designed to preserve privacy, adding linkages to clinical databases may require a re-structure even with de-identification protocols in place (Bauer et al., 2020; Massive coronavirus sequencing efforts urgently need patient data - Nature India, 2020). For example, in regions with low prevalence, the exact location in combination with height and weight can be identifiable. For such a future addition, a clinical record guardian may be needed to provide access to clinical data via a tier system.

Other likely factors are the time-consuming aspect of a task that does not immediately save lives, compounded by the reference laboratories having to chase up busy clinical teams who may not see the immediate benefit. While compiling patient information will remain a labour-intensive task, at least the design of the input forms can help by not increasing the data-entry burden unduly.

Walking the fine line between capturing enough data in a standardized way, but also making entry not so onerous to deter individuals from wanting to submit information in the first place, is an ongoing challenge. For our case-study FHIR IG, we have chosen to make most of the data field simple check boxes, with the possibility of selecting more granular concepts using auto-complete style search powered by the terminology server. This expands on the recommendations from the WHO's guidance, while still ensuring quick and efficient data capture with consistency across the world. These

high-level categories should be revisited regularly to incorporate any novel signs and symptoms that are identified as being associated with the infection.

Implementing the COVID-19 symptom capture as check boxes is possible because most guidelines provide a limited list of symptoms to capture. Should this list be expanded in the future or for other viruses, such as influenza virus and respiratory syncytial virus, 'auto-complete' search or drop-down list can be easily added to the FHIR IG.

However, it must be stressed that manual data re-entry even with the use of a FHIR questionnaire can only be an intermediate solution as efficacy and accuracy can only be achieved by enabling interoperability with clinical systems and data pre-population through FHIR standards like 'Structured Data Capture'. For example, while McAuley et al. (2020) were investigating the D614G mutation (Korber et al., 2020), it was discovered that VIC31 and VIC50 isolates originate from the same patient, and it is likely that more such duplicates exist and complicate data analysis.

Data consistency issues will be an even greater challenge for low-resource and developing countries. As outlined by Banu et al., efficient contact tracing is crucial as a single cluster can rapidly spread in densely populated countries such as India (Banu et al., 2020). This is currently hampered by a lack of detailed reporting in India such as the patient's home state being different to that of the submitting laboratory, which can confuse epidemiological analyses, as was shown to be the case recently (Mehrotra et al., 2020).

4 | RECOMMENDATIONS

In order to assess and detect a shift in the clinical presentation of COVID-19, de-identified patient data need to be collected in a more systematic way. We hence recommend three elements for the medical and scientific community to consider for capturing COVID-19 better:

1. Define the common information model and standard code sets to describe patient 'journeys' in coordination with the WHO.
2. Work towards full interoperability where the EMRs can pre-populate the FHIR questionnaire; however, this first step of creating a standard questionnaire with FHIR IG (Metke-Jimenez & Hansen, 2019) already represents a substantial advancement.
3. Update clinical workflows to revisit entries and update information.

Anticipating the opportunity for retrospective data intake in a more controlled fashion, GISAID has a mechanism to reach out to data submitters to update entries. As a more immediate improvement, GISIAD now provides a filter for serving out cleaned data correcting and consolidating 26,838 entries (see consolidated entries as of 15th May 2020 in Table S2), which is aided by a data curation tool.

These measures are valuable because the pandemic could well continue/re-emerge for some time creating the potential for new virus strains to be linked to decreased or increased case severity and/or fatality, and potentially affect the efficacy of vaccines and countermeasures. GISAID does offer clade/lineage and variant information to facilitate genotype–phenotype analyses. Gaining experience in controlled data collection increases our preparedness for future 'Disease X' outbreaks and pandemics, and enables the better support of research work for other infectious diseases such as influenza and the respiratory syncytial virus.

ACKNOWLEDGEMENTS

ST is supported by a grant awarded to Timothy Barkham and Swaine Chen by the Temasek Foundation and by the Genome Institute of Singapore; ST and SMS are supported by the Agency for Science, Technology and Research (A*STAR). AP's work on the automated meta-data curation tool is supported by Institut Pasteur with feedback from its EpiCoV™ data curation team aiding GISAID. SSV acknowledges grant funding from the Coalition for Epidemic Preparedness Innovations (CEPI).

CONFLICTS OF INTERESTS

The authors declare that there are no competing interests.

AUTHOR CONTRIBUTION

DCB, SSV and DPH conceived the paper. ST and AP structured the data. AM-J, LOWW and YJ conducted the analysis. DCB, SM-S, KE, DPH and SSV wrote the paper. All authors reviewed and finalized the document.

ETHICAL APPROVAL

Not applicable.

DATA AVAILABILITY STATEMENT

Not applicable.

ORCID

Denis C. Bauer https://orcid.org/0000-0001-8033-9810
Seshadri S. Vasan https://orcid.org/0000-0002-7326-3210

REFERENCES

Banu, S., Jolly, B., Mukherjee, P., Singh, P., Khan, S., Zaveri, L., & Sowpati, D. T. (2020). A distinct phylogenetic cluster of Indian SARS-CoV-2 isolates. *Open Forum Infectious Diseases*, *7*(11), https://doi.org/10.1093/ofid/ofaa434

Bauer, D. C., Tay, A. P., Wilson, L. O. W., Reti, D., Hosking, C., McAuley, A. J., Vasan, S. S. (2020). Supporting pandemic response using genomics and bioinformatics: a case study on the emergent SARS-CoV-2 outbreak. *Transboundary and Emerging Diseases*, *67*(4), 1453–1462. https://doi.org/10.1111/tbed.13588.

Chu, H., Chan, J.-F.-W., Yuen, T.-T.-T., Shuai, H., Yuan, S., Wang, Y., & Yuen, K.-Y. (2020). Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies

of COVID-19: An observational study. *The Lancet Microbe*, *1*(1), E14–E23. https://doi.org/10.1016/S2666-5247(20)30004-5

Dyer, C. (2020). Covid-19: Rules on sharing confidential patient information are relaxed in England. *BMJ (Clinical Research Ed.)*, *369*, m1378. https://doi.org/10.1136/bmj.m1378

Elbe, S., & Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*, *1*(1), 33–46. https://doi.org/10.1002/gch2.1018

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L. & REDCap Consortium (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W. & Sheffield COVID-19 Genomics Group (2020). Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *Cell*, *182*(4), https://doi.org/10.1016/j.cell.2020.06.043

Massive coronavirus sequencing efforts urgently need patient data - Nature India (2020). Retrieved from https://www.natureasia.com/en/nindia/article/10.1038/nindia.2020.75. accessed May 27, 2020

McAuley, A. J., Kuiper, M. J., Durr, P. A., Bruce, M. P., Barr, J., Todd, S., & Vasan, S. S. (2020). Experimental and in silico evidence suggests vaccines are unlikely to be affected by D614G mutation in SARS-CoV-2 spike protein. *Npj Vaccines*, *5*(1), 96. https://doi.org/10.1038/s41541-020-00246-8

Mehrotra, K. (2020). 'Unassigned' coronavirus cases near 3,000, rise as curbs on movement lifted. The Indian Express. https://indianexpress.com/article/india/coronavirus-india-lockdown-unassigned-cases-near-3000-rise-as-curbs-on-movement-lifted-6428789/

Menni, C., Valdes, A. M., Freidin, M. B., Sudre, C. H., Nguyen, L. H., Drew, D. A., Ganesh, S., Varsavsky, T., Cardoso, M. J., El-Sayed Moustafa, J. S., Visconti, A., Hysi, P., Bowyer, R. C. E., Mangino, M., Falchi, M., Wolf, J., Ourselin, S., Chan, A. T., Steves, C. J., & Spector, T. D. (2020). Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine*, *26*(7), 1037–1040.

Metke-Jimenez, A., & Hansen, D. (2019). FHIRCap: Transforming REDCap forms into FHIR resources. *AMIA Joint Summits on Translational Science Proceedings AMIA Summit on Translational Science*, *2019*, 54–63.

Metke-Jimenez, A., Steel, J., Hansen, D., & Lawley, M. (2018). Ontoserver: A syndicated terminology server. *Journal of Biomedical Semantics*, *9*(1), 24. https://doi.org/10.1186/s13326-018-0191-z

Shrimp browser citable link for COVID-19 symptoms (n.d.). Retrieved from https://ontoserver.csiro.au/shrimp/vs.html?system=undefined&valueSetUri=http%3A%2F%2Fgenomics.ontoserver.csiro.au%2Ffhir%2Fcovid19%2FValueSet%2FCovid19SymptomsValueSet&valueSetId=Covid19SymptomsValueSet&fhir=https://r4.ontoserver.csiro.au/fhir. accessed May 12, 2020

Su, Y., Anderson, D., Young, B., Zhu, F., Linster, M., Kalimuddin, S., & Smith, G. (2020). Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2. *mBio*, *11*(4), https://mbio.asm.org/content/11/4/e01610-20

The COVID-19 Host Genetics Initiative (2020). The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics*, *28*, 715–718. https://www.nature.com/articles/s41431-020-0636-6

WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection (2020). A minimal common outcome measure set for COVID-19 clinical research. *The Lancet Infectious Diseases*, *20*(8), e192–e197. https://doi.org/10.1016/S1473-3099(20)30483-7

Worldometers coronavirus. (n.d.). Retrieved from https://www.worldometers.info/coronavirus/. accessed April 28, 2020

Yao, H., Lu, X., Chen, Q., Xu, K., Chen, Y., Cheng, M., Chen, K., Cheng, L., Weng, T., Shi, D., Liu, F., Wu, Z., Xie, M., Wu, H., Jin, C., Zheng, M., Wu, N., Jiang, C., & Li, L. (2020). Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and with clinical implications in vivo. *Cell Discovery*, *6*(76), https://www.nature.com/articles/s41421-020-00226-1

Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., Jia, X., Wu, M., Shi, B., Xu, S., Chen, J., Wang, W., Chen, B., Jiang, L. U., Yu, S., Lu, J., Wang, J., Xu, M., Yuan, Z., … Lu, H. (2020). Viral and host factors related to the clinical outcome of COVID-19. *Nature*, *583*(7816), 437–440. https://doi.org/10.1038/s41586-020-2355-0

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Bauer DC , Metke-Jimenez A , Maurer-Stroh S , et al. Interoperable medical data: The missing link for understanding COVID-19. *Transbound Emerg Dis*. 2021;00:1–8. https://doi.org/10.1111/tbed.13892