response alongside the political response across countries," she says.

This collection of studies is a mix of existing population cohorts and studies established early in the pandemic. Existing cohorts are advantageous because their compositions tend to reflect that of the population overall, so their results can be generalized. And because long-running population cohorts will have data on participants from before the pandemic, they can quantify changes in mental health accurately, says epidemiologist Klaus Berger at the University of Muenster in Germany, who chairs the German National Cohort, one of the world's largest health cohorts.

But large, established cohorts move relatively slowly and sample infrequently. The newer cohorts lack the baseline of data collected before the pandemic, but many can follow the dynamics of the crisis in a nimbler way.

Fancourt leads one of the largest new studies, the UK COVID-19 Social Study. The study recruited — mostly through social media — more than 72,000 UK adults in the first few weeks of the country's first lockdown, in March. Participants fill in a weekly ten-minute online questionnaire, which includes questions that identify feelings of anxiety or depression.

Real-time data

"With survey responses coming in at a rate of one every 20 seconds, we get information about how people are responding psychologically and socially to the pandemic in real time, and see specifically how it's changed in response to things like new government measures coming in, or lockdown measures being eased," says Fancourt. For example, she says,

> ## "We have a natural experiment in how different policies impact people's mental health."

the high levels of anxiety and depression the study found in its early weeks reduced during the lockdown, rather than increased as some had anticipated.

"Together, these types of study will tell us how government policies are experienced across different segments of societies and will help us understand how we should manage this pandemic, and future pandemics," says Nazroo, who is participating in the European Union-wide Survey on Health, Ageing and Retirement in Europe cohort and other surveys related to COVID and mental health.

Another study, called the COVID-19 Health Care Workers Study, aims to quantify how health workers, who have faced unprecedented levels of illness and death, have coped. The study is collecting data in 21 countries, including low-income nations in Latin America and Africa where mental-health resources are very limited. "We want to compare across countries to know what is happening that is different," says Olatunde Ayinde, a researcher on the study's Nigerian arm. He thinks that geographical variations are likely to stem from differences in the quality of mental-health services, the availability and types of social care on offer and poverty levels. Many countries in Africa have just a fraction of the mental-health practitioners that high-income countries have. "We want to know what is responsible for the differences," says Ayinde.

Additional reporting by Paul Adepoju.

SCIENTISTS CALL FOR OPEN SHARING OF PANDEMIC GENOME DATA

But others say that certain restrictions encourage faster sharing.

By Richard Van Noorden

Hundreds of scientists are urging that SARS-CoV-2 genome data should be shared more openly to help analyse how viral variants are spreading around the world.

Researchers have posted huge numbers of SARS-CoV-2 genome sequences online since January 2020. The most popular data-sharing platform, called GISAID, now hosts more than 450,000 viral genomes; Soumya Swaminathan, the chief scientist at the World Health Organization (WHO), has called it a game-changer in the pandemic. But it doesn't allow sequences to be reshared publicly, which is hampering efforts to understand the coronavirus and the rapid rise of new variants, argues Rolf Apweiler, co-director of the European Bioinformatics Institute (EBI) near Cambridge, UK, which hosts its own large genome database that includes SARS-CoV-2 sequences.

In a letter released on 29 January (see go.nature.com/3rtjgj5), Apweiler and others call for researchers to post their genome data in one of a triad of databases that don't place any restrictions on data redistribution: the US GenBank, the EBI's European Nucleotide Archive (ENA) and the DNA Data Bank of Japan, which are collectively known as the International Nucleotide Sequence Database Collaboration (INSDC).

Anyone can anonymously access the INSDC's data and use them as they want, but GISAID requires that users confirm their identity and agree not to republish the site's genomes without permission from the data provider. This means that studies building on GISAID data — such as those that create evolutionary trees analysing how SARS-CoV-2 variants are related — can't publish full data so that others can easily check their analyses or further build on their data set. They must instead direct readers back to the GISAID site.

The letter says the scientific community should "remove barriers that restrain effective data sharing", but doesn't mention GISAID specifically. It is signed by more than 500 scientists, including the 2020 chemistry Nobel laureate Emmanuelle Charpentier, and the head of the COVID-19 Genomics UK Consortium, Sharon Peacock. Where scientists have already established submissions to other databases, the letter states, "these submissions should continue in parallel".

But many researchers who work with GISAID say that its terms of access are a benefit, because they encourage hesitant researchers to share data online speedily, without fear that others will use the results without credit. "The reason so many labs have provided SARS-CoV-2 genomes to GISAID is precisely because of the data-access agreement that restricts public resharing," says Sebastian Maurer-Stroh, a bioinformatician at Singapore's Agency for Science, Technology and Research. GISAID has worked with many labs to assist them to share data, he says.

GISAID stands for the Global Initiative on Sharing Avian Influenza Data; an international consortium of researchers helped to set it up as a non-profit foundation in 2008, to address researchers' reluctance to share data on influenza strains. Some nations, including Indonesia, a hotspot for avian flu, feared that pharmaceutical firms would create drugs and vaccines using the sequence data without crediting the original data providers or sharing the benefits of the work with them. But they were persuaded to share sequences rapidly on GISAID; in March 2013, for instance, China published sequences of H7N9 avian flu in the database on the same day it informed the WHO of three infections in people. "GISAID

encourages and incentivizes real-time data sharing by parties who would otherwise be reluctant to share, by ensuring that they retain their rights in their data," says a spokesperson for the initiative.

"This issue is not only about science, but also about sovereignty and equity," says Marie-Paule Kieny, a vaccine researcher at INSERM, the French national health-research institute in Paris. "GISAID empowers the rapid flow of SARS-CoV-2 sequence data with maximal impact," she says, because scientists depositing sequences can trust that their rights will be respected by data users.

Senjuti Saha, a microbiologist who works on SARS-CoV-2 genomes at the Child Health Research Foundation in Dhaka, says that she appreciates the call for open data beyond what GISAID offers, but worries that it might further dissuade researchers in low- and middle-income countries (LMICs) from uploading data until they have analysed them. During the pandemic, she says, some LMICs have started doing more viral sequencing, although labs often lack computational infrastructure. She says that she's seen LMIC coronavirus data taken out of context by academics in wealthier countries who don't consult or credit the data providers. "We really want to share our

> ### "We really want to share our data, but it is heart-breaking and demotivating when we don't get the credit."

data, but it is heart-breaking and demotivating when we know we worked so hard to generate data, but we don't get the credit for it," she says.

The letter, says Kieny, "seems to me like an initiative from European and high-income countries not fully informed on the critical need to ensure that low-resource countries accept to share sequences freely, so that the public-health impact of sequencing of pathogens such as SARS-CoV-2 is maximized".

ENA head Guy Cochrane says the EBI is aware of the global issues around data and benefit sharing, and is actively involved in finding benefit-sharing mechanisms that empower countries in the global south and keep data open. But even well-resourced European countries could do more to share their data openly, he says.

Some researchers told Nature that besides arguments about equity and openness, there is an issue with GISAID's differential control over how registered users can download its data. Some users must download files in small batches, for instance, but others can get an entire data set in bulk with GISAID approval. The GISAID spokesperson says that's because the initiative needs to know who is using its data and for what reason, so that nothing is erroneously redistributed.

AI MATHS WHIZ CREATES TOUGH NEW PROBLEMS FOR HUMANS TO SOLVE

Algorithm named after Srinivasa Ramanujan suggests formulae, some of which are difficult to prove.

By Davide Castelvecchi

Researchers have built an artificial intelligence (AI) that can generate new mathematical formulae — including some problems that continue to challenge mathematicians.

The Ramanujan Machine is designed to generate new ways of calculating the digits of important mathematical constants, such as π or e, many of which are irrational, meaning that they have an infinite number of non-repeating decimals.

The AI starts with well-known formulae to calculate the digits — the first few thousand digits of π, for example. From those, the algorithm tries to predict a new formula that does the same calculation just as well. The process produces a good guess called a conjecture — it is then up to human mathematicians to prove that the formula can correctly calculate the whole number.

The project team began to make the conjectures public on its website in 2019 (see go.nature.com/3td0ky3), and researchers have since proved several of them correct. But some remain open questions, including one on Apery's constant, which has important applications in physics. "The last result, the most exciting one, no one knows how to prove," says physicist Ido Kaminer, who leads the project at the Technion — Israel Institute of Technology in Haifa. The automated creation of conjectures could point mathematicians towards connections between branches of maths that people did not know existed, he adds.

The project — described in Nature on 3 February (G. Raayoni et al. Nature **590**, 67–73; 2021) — is named after Srinivasa Ramanujan, an Indian mathematician who was active in the early twentieth century. Ramanujan rarely wrote the types of proof that appear in conventional maths papers. Instead, he filled entire notebooks with formulae that he believed came from a goddess who appeared in his dreams. His work continues to inspire new research long after his death, aged 32, in 1920.

The Ramanujan Machine currently has limited applications: so far, the algorithms can generate only formulae of a particular type, called continued fractions. These express a number as an infinite sequence of fractions nested in each other's denominators.

Mathematician Srinivasa Ramanujan.

Kaminer's team has experimented with a range of algorithms for finding continued fractions, and applied them to various conceptually important numbers. One of them is Catalan's constant, a number that originated from nineteenth-century Belgian mathematician Eugène Catalan's studies.

Catalan's constant is approximately 0.916, but it is so mysterious that no one has yet worked out whether it is rational — that is, whether it can be expressed as a fraction of two whole numbers. The best that mathematicians have been able to do is prove that its 'irrationality exponent' — a measure of how hard it is to approximate a number using rationals — is at least 0.554. Proving that Catalan's constant is irrational would be equivalent to proving that its irrationality exponent is greater than 1. Formulae generated by the Ramanujan Machine have enabled Kaminer's team to improve slightly on the best human result, bringing the exponent up to 0.567.

"The fact that they have improved the irrationality exponent ... reveals that they are able to make contributions to really hard problems," says George Andrews, a mathematician at Pennsylvania State University in University Park. But the contributions made so far are not of the calibre that using Ramanujan's name would suggest, he says. "Calling this the Ramanujan Machine is over the top."

Kaminer's team plans to broaden the AI's technique so that it can generate other kinds of mathematical formula.