



Why open science is critical to combatting COVID-19

Updated 12 May 2020

Key messages

- In global emergencies like the coronavirus (COVID-19) pandemic, open science policies can remove obstacles to the free flow of research data and ideas, and thus accelerate the pace of research critical to combating the disease.
- While global sharing and collaboration of research data has reached unprecedented levels, challenges remain. Trust in at least some of the data is relatively low, and outstanding issues include the lack of specific standards, co-ordination and interoperability, as well as data quality and interpretation.
- To strengthen the contribution of open science to the COVID-19 response, policy makers need to ensure adequate data governance models, interoperable standards, sustainable data sharing agreements involving public sector, private sector and civil society, incentives for researchers, sustainable infrastructures, human and institutional capabilities and mechanisms for access to data across borders.

In the current global emergency, scientific discovery has evolved much more rapidly than before. The full genome of COVID-19 was published barely a month after the first patient was admitted into Wuhan hospital, as [an open-access publication in *The Lancet*](#). This is to be compared with a five-month delay in the case of SARS outbreak in 2002-03, a large part of this delay being due to an information blackout in the first months of the SARS epidemic.

Lessons from previous outbreaks have underscored the importance of sharing data and publications in order to combat the disease. Key enablers of this sharing are:

WHY OPEN SCIENCE IS CRITICAL TO COMBATTING COVID-19 © OECD 2020



- building and maintaining trust between parties sharing research data
- reciprocity of research data sharing
- inclusive inter-sectoral collaboration, based on pre-defined roles and responsibilities
- creation of a preparedness and response system that fits all emerging infectious diseases with appropriate supportive technical infrastructure as well as pre-defined access rights and responsibilities of stakeholders
- having trusted international collaborating partners as external advisor and reference centres
- addressing the barriers to sharing of research data, with solutions that take into account the complexity and multitude of root causes that cause these barriers.

This brief provides an overview of achievements in sharing data, publications, and creating online collaborative platforms, and outlines the remaining challenges. It concludes by providing a roadmap towards even better and more resilient policies for the future.

Achievements of open science initiatives and commitments

In January 2020, 117 organisations – including journals, funding bodies, and centres for disease prevention – signed a statement titled “[Sharing research data and findings relevant to the novel coronavirus outbreak](#)”, committing to provide immediate open access for peer-reviewed publications at least for the duration of the outbreak, to make research findings available via preprint servers, and to share results immediately with the World Health Organization (WHO). This was followed in March by the [Public Health Emergency COVID-19 Initiative](#), launched by 12 countries¹ at the level of chief science advisors or equivalent, calling for open access to publications and machine-readable access to data related to COVID-19, which resulted in [an even stronger commitment](#) by publishers.

[The Open COVID Pledge](#) was launched in April 2020 by an international coalition of scientists, lawyers, and technology companies, and calls on authors to make all intellectual property (IP) under their control available, free of charge, and without encumbrances to help end the COVID-19 pandemic, and reduce the impact of the disease. Some notable signatories include Intel, Facebook, Amazon, IBM, Sandia National Laboratories, Hewlett Packard, Microsoft, Uber, Open Knowledge Foundation, the Massachusetts Institute of Technology, and AT&T. The signatories will offer a specific non-exclusive royalty-free Open COVID license to use IP for the purpose of diagnosing, preventing and treating COVID-19.

Following these commitments, a number of leading publishers and journals² are providing open access, and numerous data servers are available to share epidemiological, clinical and genomics data. Data, protocols and standards used to collect the data are also being shared. The [CORD-19](#) (COVID-19 Open Research Dataset), contains 57 000 entries, including 41 000 full-text machine-readable articles on COVID-19 and related coronaviruses, and serves as a basis for data mining by machine learning techniques, in order to answer a set of open questions about COVID-19.

Also illustrating the power of open science, online platforms are increasingly facilitating collaborative work of COVID-19 researchers around the world. A few examples include:

- [Nextstrain](#) and [Gisaid](#) allow tracking the epidemic spread through genetic mutations.
- Modelling of the epidemic spread is enabled by platforms such as [MOBS Lab](#) or [MIDAS](#).
- Research on treatments and vaccines is supported by [Elixir](#), [REACTing](#), [CEPI](#) and others.
- Crowdsourcing efforts like [Foldit](#) involve challenges, while hackathons have emerged including [#EUvsVirus](#) and [COVID-19 virtual study-a-thon](#).

¹ Australia, Brazil, Canada, Germany, India, Italy, Japan, New Zealand, Korea, Singapore, the United Kingdom and the United States of America.

² The British Medical Journal, *The Lancet*, *Nature*, Elsevier, Springer, Cambridge University Press, Wiley and others.



- [Vivli](#) is a platform that offers an easy way to request anonymised data from completed clinical trials.
- The European Commission and several partners established a [COVID-19 data portal](#) in April 2020 to enable rapid and open sharing of research data to advance research on the disease.
- Computing resources are being offered by the [COVID-19 High Performance Computing Platform](#) and [Folding@home](#), a distributed computing platform, is providing more than 1.5 exaFlops.

Remaining challenges

While **clinical, epidemiological and laboratory data about COVID-19** is widely available, including genomic sequencing of the pathogen, a number of challenges remain:

- All data is not sufficiently findable, accessible, interoperable and reusable (FAIR), or not yet FAIR data.
- Sources of data tend to be dispersed, even though many pooling initiatives are under way, curation needs to be operated “on the fly”.
- Providing access to personal health record sharing needs to be readily accessible, pending the patient’s consent. Legislation aimed at fostering interoperability and avoiding information blocking are yet to be passed in many OECD countries. Access across borders is even more difficult under current data protection frameworks in most OECD countries.
- In order to achieve the dual objectives of respecting privacy while ensuring access to machine readable, interoperable and reusable clinical data, [the Virus Outbreak Data Network \(VODAN\)](#) proposes to create FAIR data repositories which could be used by incoming algorithms (virtual machines) to ask specific research questions.
- In addition, many issues arise around the interpretation of data – this can be illustrated by the widely followed epidemiological statistics. Typically, the statistics concern “confirmed cases”, “deaths” and “recoveries”. Each of these items seem to be treated differently in different countries, and are sometimes subject to methodological changes within the same country.
- Specific standards for COVID-19 data therefore need to be established, and this is one of the priorities of the [UK COVID-19 Strategy](#). A working group within Research Data Alliance has been set up to propose such standards at an international level.
- In some cases it could be inferred that the transparency of the statistics may have guided governments to restrict testing in order to limit the number of “confirmed cases” and avoid the rapid rise of numbers. Lower testing rates can in turn reduce the efficiency of quarantine measures, lowering the overall efficiency of combating the disease.

Concerning **open access to publications**, challenges also remain:

- The current positive engagements by publishers are set to expire in three months, and the sustainability in the long run is uncertain. It also concerns a small core of knowledge directly linked to COVID-19, and fails to open up the broader interdisciplinary³ knowledge base needed for full understanding of the virus. A recent [study](#) shows that less than one-third of the interdisciplinary publications referenced in COVID-19 are open-access.
- It remains to be seen how the crisis will impact the broader discussion about progress towards open-access publishing, including initiatives such as [Plan S](#), an international project that requires all scientific publications resulting from research funded by public grants be available in open access.⁴

³ Relevant fields concern 138 fields, including virology, biochemistry and molecular biology, immunology, general biomedical research, microbiology, medicine, pharmacology, cellular biology, genetics, naturology, respiratory system and public health.

⁴ Some signatories of the Wellcome Trust statement cited above are opponents of Plan S.



- Preprints⁵ have been encouraged as a vehicle for rapid knowledge diffusion during the crisis, and this has largely proved positive. While preprint circulation allows for increased speed of diffusion, it presents risks of quality control. For example, a paper published on the BioRxiv server on 2 February erroneously asserted that the COVID-19 virus sequence might have been man-made. Luckily, the error was quickly spotted by fellow scientists and the paper was removed within hours.

Challenges also remain with **platforms that are springing up to facilitate research collaboration**:

- Communication and co-ordination between the multiple initiatives need to be improved. In some cases, structuring within a hub and spoke network could help to enhance usability.
- The lack of co-ordination is compounded by issues of interoperability. Different platforms have different architectures and it is essential to tackle this during the initial phase of the response.
- The target audiences of the different platforms are sometimes unclear. They may include researchers, clinicians, policy makers and/or the general public. The needs of the three audiences need to be clarified and catered to.
- Finally, the sustainability of the platforms for research collaboration is not a given. Funding is available in the short term as a crisis response measure, but may not be guaranteed in the long term as other priorities emerge.

The way forward: Resilient open science policies

Given the achievements and challenges of open science in the current crisis, lessons from prior experience in OECD countries can be drawn to assist the design of open science initiatives to address the COVID-19 crisis. Applying the general framework of recommendations referenced in prior OECD work on access to data from publicly funded science, technology and innovation, the following actions can help to further strengthen open science in support of responses to the COVID-19 crisis:

1. **Developing of data governance models that allow for open research data by default**, while preserving individual privacy. This involves setting up strong consent mechanisms monitored by ethical review boards. Ethical frameworks are needed that protect all parties (e.g. patients, healthcare workers, institutions) from immediate and longer-term consequences.
2. **Providing regulatory frameworks that would enable interoperability within the networks of large electronic health records providers**, patient mediated exchanges, and peer-to-peer direct exchanges.⁶ Data standards need to ensure that data is findable, accessible, interoperable and reusable, including general data standards, as well as specific standards for the pandemic. The Research Data Alliance has set up a COVID-19 working group that is set to provide recommendations on this aspect in April 2020.
3. **Working together** by public actors, private actors, and civil society to **develop and/or clarify a governance framework for the trusted reuse of privately-held research data** toward the public interest. This framework should include governance principles, open data policies, trusted data reuse agreements, transparency requirements and safeguards, and accountability mechanisms, including ethical councils, that clearly define duties of care for data accessed in emergency contexts.
4. **Clarifying incentives and rewards for researchers, and require the immediate disclosure of data, software and protocols for publication**. Institutional and national policies should address issues of recognition and cultural/structural barriers among data contributors, shifting the culture to one where sharing is the norm.

⁵ Preprints are drafts of publications submitted to scientific journals, and awaiting peer review.

⁶ For instance, new rules by the US Office of the National Coordinator for Health IT and Centers for Medicare & Medicaid Services now require providers to adopt application programming interfaces to allow patients easily access their medical data at no cost.



5. **Securing adequate infrastructure (including data and software repositories, computational infrastructure, and digital collaboration platforms)** to allow for recurrent occurrences of emergency situations. This includes a global network of certified trustworthy and interlinked repositories with compatible standards to guarantee the long-term preservation of FAIR COVID-19 data, as well as the preparedness for any future emergencies.
6. **Ensuring that adequate human capital and institutional capabilities are in place to manage, create, curate and reuse research data** – both in individual institutions and in institutions that act as data aggregators, whose role is real-time curation of data from different sources.
7. **Enabling access to sensitive research data across borders on a more restricted basis in secure environments.** This primarily concerns clinical data which may not be allowed to leave the original repository, but could potentially be accessed by mobile⁷ algorithms which could use the data to answer specific research questions.

Further reading

- OECD (2020), *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*, OECD Publishing, Paris, <https://doi.org/10.1787/947717bc-en>.
- OECD (2020), “Open science initiatives related to the COVID-19 pandemic”, webpage, OECD, Paris, <https://community.oecd.org/docs/DOC-172520>.
- OECD (2020), “Ensuring data privacy as we battle COVID-19”, OECD, Paris, <https://www.oecd.org/coronavirus/policy-responses/ensuring-data-privacy-as-we-battle-covid-19/>.
- OECD (2017), “Business models for sustainable research data repositories”, *OECD Science, Technology and Industry Policy Papers*, Vol. 47, OECD Publishing, Paris, <https://doi.org/10.1787/302b12bb-en>.
- OECD (2016), “Research ethics and new forms of data for social and economic research”, *OECD Science, Technology and Industry Papers*, Vol. 34, OECD Publishing, Paris, <http://dx.doi.org/doi.org/10.1787/23074957>.
- OECD (2015), “Making open science a reality”, *OECD Science, Technology and Industry Policy Papers*, Vol. 25, OECD Publishing, Paris, <http://dx.doi.org/10.1787/5jrs2f963zs1-en>.
- OECD (2006), *Recommendation of the Council concerning Access to Research Data from Public Funding*, OECD, Paris, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>.

This paper is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.

www.oecd.org/sti – sti_contact@oecd.org –  [@OECDInnovation](https://twitter.com/OECDInnovation) – <http://oe.cd/stinews>

⁷ A mobile algorithm is one which is sent over the Internet to access a remote data set. The idea is that data does not move, virtually or otherwise. It is just accessed by an algorithm which is sent to analyse it.

