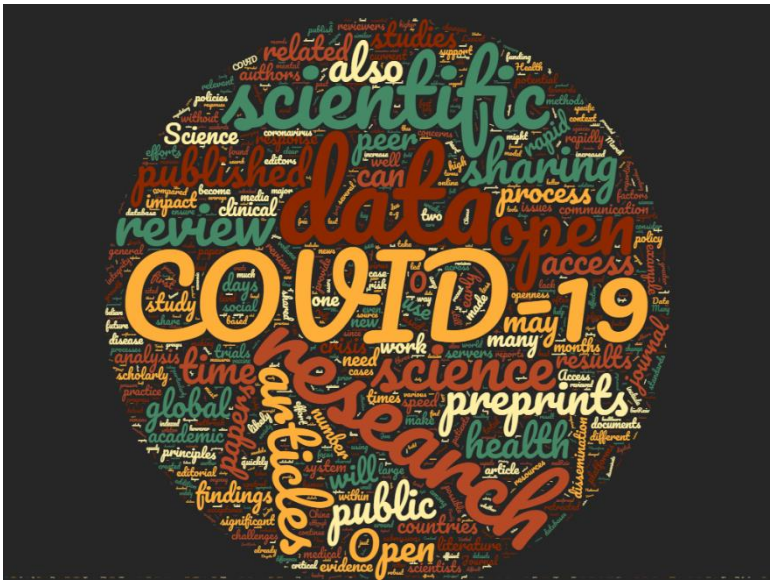


1/22/2024

Open Science and COVID Research

Final Report

Daniel Spichtinger

This study was commissioned by frontiers.

Table of Content

Key Messages and Executive Summary 4

1. Introduction 6

1.1. About the consultant 6

1.2. Objectives and Scope 6

2. Results of the Literature Review 7

2.1. Regular Google Search Findings 8

2.2. Google Scholar Search Findings 9

2.3. Specific findings related to open access 10

a) bibliometrics: largely open, but 10

b) Volume of (open access) literature: infodemic 11

c) Key factors for speed: time to acceptance/publication and pre-prints 12

2.4. Specific findings related to Open Data and Data sharing 15

2.5. Resources, tools, and national case studies 18

2.6. Framework conditions and other open science components 18

2.7. Conclusions 20

3. Survey Results 21

3.1. General factors and policies 21

3.2. Access to scientific publications and related issues 22

3.3. Access to scientific data and related issues 25

3.4. Other open science practices 29

3.5. Conclusions 32

4. Interview results 33

4.1. Feedback on the main factors identified in the literature for the speed of COVID research 34

4.2. Feedback on access to scientific publications and speed of COVID research 35

4.3. Feedback on quality concerns, pre-prints and peer review 36
4.4. Feedback on access to research data and speed of COVID research 37
4.5. Feedback on data quality 38
4.6. Feedback on importance of access to publications vis-à-vis access to data 40
4.7. Feedback on other open science issues and the speed of COVID research 41
4.8. Feedback on the effectiveness of the scientific system in tackling COVID 42
4.9. Final thoughts on open science and the speed of COVID research 43
4.10. Conclusions from the interviews 43
5. Overall Study Conclusions and Policy Implications 47
6. Annexes 49

Key Messages and Executive Summary

Key Messages

i. Access to scientific publications was a very important factor contributing to the speed of COVID research. Although the majority of researchers was satisfied with the level of open access to COVID publications, some limitations remain.
ii. Open data was also considered important for the speed of COVID-related research but less so than open access to publications. Only a narrow majority was satisfied with the quality of available COVID data, and many pointed out that there was room for improvement in terms of accessibility, quality, and legal frameworks for data sharing across borders.
iii. The value of pre-prints, the current state of peer-review, and the role of speed versus quality in scientific research were seen as closely connected to open access and the speed of COVID research and were controversially discussed. The lack of consensus shows a scientific system still in transition towards open science.
iv. The actions to open up knowledge taken during the pandemic (such as the Joint Statement) have delivered a boost towards achieving open science but should not be considered a silver bullet. In the post COVID era we need to guard against backsliding and continue to make progress by addressing remaining challenges. Ultimately, implementing open science is a marathon, not a sprint.

Open science, in particular open access scientific publications and data sharing, have sometimes been depicted as key assets in combatting the COVID-19 pandemic, in particular regarding the unprecedented speed at which vaccines have been developed. This report provides empirical evidence to explore whether and to what extent open science practices, most notably open access to publications and open research data have influenced the speed of COVID research. For this purpose, a literature review and a survey of more than 200 COVID researchers (phase 1) was conducted, followed by qualitative interviews with six scientists (phase 2).

In phase 1, our **literature review** found that non-scientific documents (e.g. from funders) explained the speed of COVID research primarily through factors such as previous research (e.g. on SARS), the unprecedented level of funding, worldwide collaboration, new technology, and regulatory and management innovation. Open science could be considered an implicit component of the worldwide collaboration category. When explicitly adding "open science" and its subcomponents to the search query, there were hits from governmental agencies that stress the importance of open science during the pandemic. Analysing scientific peer reviewed publications (N=112) open science components (open access /data sharing) were more explicitly named as key factors of the scientific response to COVID-19 and were often discussed co-dependently with the increased speed of scientific production and the sheer number of publications and data produced. These topics were controversially debated, revealing tensions between progressive (sometimes radical) members of the scientific community, who consider any shortcomings to be the fault of too little and not too much open science, and conservatives, who would like to retain or expand the role of traditional gatekeeps as a bastion for quality.

Based on the results of the literature review a **survey** of researchers which have published about COVID with Frontiers (n=208) was then conducted.

Summary of Key Survey Results

- Importance of access to scientific publications for the speed of COVID research: 4,19 out of 5[1] (89% provided 4 or 5 star rating)
- Percentage of COVID publications that respondents said were open: 72% (average)
- Percentage of respondents satisfied with access to scientific publications: 82%
- Importance of access to data for the speed of COVID research: 3,88 out of 5 (68% provided 4 or 5 star rating)
- Percentage of datasets that respondents said were open: 58%
- Percentage of respondents satisfied with access to data: 64%
- Percentage of respondents satisfied with data quality: 53%

Further results of the survey indicate that the participants considered the measures to speed up peer review of COVID literature as positive (62%) whilst at the same time considering that there is a further need to reform peer review (73%). Respondents were split on whether speeding up of COVID research has had a negative effect on the quality of research: 40% strongly agreed or agreed with this statement but 35% are undecided ("neither agree or disagree") and 25% disagree or disagree strongly. Last but not least a super majority of respondents (93%) agreed that scientific information produced by public funding should be available as openly as possible. They also agree that open **science practices beyond open access to publications and data sharing are necessary,** with 79% agreeing that the way we assess researchers and research projects needs to change.

In the second phase of the study, **qualitative interviews** with six stakeholders involved in COVID research were conducted, based on the previous outcomes of the literature review and the survey. Overall, the responses suggest that, while open access publications have been critical for the development of knowledge around COVID-19 during the pandemic, there are still limitations in accessing scientific literature in some contexts, which can create inequities in the dissemination of knowledge. The interviewees had quite different opinions on the value of pre-prints, the peer-review process, and the role of speed versus quality in scientific research during the COVID-19 pandemic. However, they all acknowledge the importance of ensuring that scientific research during the pandemic is reliable and properly vetted. Most interviewees agreed that data sharing was crucial for COVID-related research, but almost all also pointed out that there was room for improvement in terms of accessibility, quality, and legal frameworks for sharing across borders. Furthermore, interviewees agreed with the need to change the culture around acknowledging data sharing and the traditional system of assessing scientific work based on the number of publications or the prestige of the journal, and two of them criticised high article processing charges. In regard to the effectiveness of the scientific system in tackling COVID, reactions were mixed: some interviewees expressed negative opinions on the handling of data and one flagged censorship during the pandemic, while others highlighted successful collaboration and sharing of

[1] 1 meaning not important at all and 5 meaning very important.

data. Regarding their final thoughts on open science and the speed of COVID research some of the interviewees provided a range of reflections, including the need to apply efforts made for COVID-19 to other diseases, the importance of speeding up funding processes, and the likelihood of pre-print services and platforms to become more popular.

The results from this study, in particular the intrinsic linkage of open access and open data to wider open science issues and their controversial discussion among the scientific community, shows a science system that is still in transition. The pandemic may have given Open Science a boost, but it is clear that open science is not yet the default setting or the modus operandi of the scientific system in the post-COVID world. The actions taken to open up knowledge during COVID (such as the Joint Statement) are not the "silver bullet" that implements open science overnight. Rather, we need to ensure that the progress that was made is built on and becomes the new baseline; vigilance against backsliding (e.g. closing publications that were made open during COVID) will also be required. Individual initiatives in different areas of open science are very necessary to give us a boost. But in the end implementing open science is a marathon, not a sprint.

1. Introduction

1.1. About the researcher

Daniel Spichtinger is an independent researcher and consultant providing strategic advice & expertise on EU research policy and Open Science. From 2012 to 2018 he was Policy Officer for Open Access to scientific publications and research data at the European Commission. After finishing his 6-year contract with the Commission, he returned to Vienna and registered as a self-employed expert for Open Science and EU Research Policy in 2018. In this capacity, he has been involved in a number of projects, such as assessing the readiness of the European Open Science Cloud in three European countries, analysing Horizon 2020 Data Management Plans and studying Open Access and Open Data policies in Azerbaijan and Malta.

1.2. Objectives and Scope

This work provides empirical evidence of a quantitative and qualitative nature to explore whether and to what extent open science practices, most notably open access to publications and open research data have been instrumental in the speed of COVID research and the development of

[2] https://creativecommons.org/licenses/by/4.0/

COVID vaccines. While this has often been alleged (e.g. Besançon et al 2021[3], OECD 2020[4], Sparc Europe[5] w.d.) there are only a few empirical studies that have looked at this issue in detail so far.

In order to do so, the study progressed in two phases:

Phase 1

(i) review and summary of the current literature

(ii) quantitative survey distributed to COVID researchers (sample largely from Frontiers authors)

Phase 2:

(iii) qualitative interviews with six COVID stakeholders on their experience with open science practices in the COVID context.

The methodology and outcomes of each phase are described in more detail in the respective sections of this report.

2. Results of the Literature Review

The literature review was conducted with both a narrower and a broader focus. The narrow focus was finding relevant material for the question on whether and to what extent open science practices (most notably open access to publications and data sharing) were an important factor in the speed with which COVID-19 vaccines have been developed. The relevant search terms for Google for this more narrowly worded focus were "speed of covid vaccination open science /open access / open data". This resulted in **22 meaningful public documents (**i.e. not scientific articles), which are analysed further below; specific references from the regular Google search engine are given in brackets ( ), with the source material being available in a supplementary excel file (see annex 1, sheet 1).

In Google *Scholar* the focus was then broadened to the key words "open science covid", "open science covid speed", "open access covid", "open access covid speed", "open data covid", "open data covid speed" "FAIR data covid", and "FAIR data covid speed". With these search results a number of promising documents were obtained; these were then followed up by researching key documents that were either directly cited in the literature or flagged up by Google Scholar as related documents. Documents that were not content related ("false positives") were removed. As a result, **112 documents were retrieved** which were then analysed (see supplementary data excel file annex 2, sheet 2). The overall majority of the items were scientific peer reviewed articles, but a number of pre-prints were also retrieved. Furthermore, a small number of items are other

[3] https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/s12874-021-01304-y.pdf
[4] https://www.oecd.org/coronavirus/policy-responses/why-open-science-is-critical-to-combatting-covid-19-cd6ab2f9/
[5] https://sparceurope.org/covid-19-and-open-science/

entities, such as blog entries or media reports. Throughout this section, items from the Google Scholar literature review are quoted with their number in square brackets [ ].

2.1. Regular Google Search Findings

The Google search term "speed of covid vaccination development" allows us to ascertain whether open science appears as a factor causing or influencing the high speed of covid vaccination development *unprompted*, that is without being explicitly included in the Google query.

A number of websites and documents outline factors that account for the speedy development of COVID-19 vaccines, some of them with the aim of reassuring the public that the vaccines are safe. Different sources sometimes put the emphasis on different components but as such the factors identified are remarkably consistent and can be summarised as follows:

- **Years of research**: previous research and knowledge about SARS (1) and MERS (3)(4)(7), which provided, as (6) puts it, "a smart head start"
- **Funding turbocharger:** reference to "Operation Warp Speed" and similar funding activities in European Union (Horizon 2020 and Horizon Europe), and other countries (1) (4)
- **Worldwide collaboration** (1), not only between scientists but also between manufacturers and distributers (5) as well as including "young biotech companies; health authorities and regulatory agencies; and a range of public and private institutions" (7); collaboration is a particular focus of (9)
- **New Technology,** most prominently mRNA (1) (4) (7) but also viral vectors (7)
- **Regulatory factors, clinical trials, parallel phasing and manufacturing:** speeding up regulatory processes (4) (7) (8) and manufacturing (7), as well as running processes in parallel, as regards both clinical trials (8) and manufacturing (i.e. starting production before approval). Parallel processes and steps are also mentioned in the Google Scholar literature review document [61]. These factors are sometimes provided together and sometimes separately; in any case they are highly interactive and dependent on each other. The quick recruitment of a large number of volunteers is also sometimes given as a factor (1), (7)

Open science and subcomponents such as open access to scientific publications or data sharing are not mentioned explicitly in these documents, although McKinsey and Company do note that "critical information on COVID-19 was quickly made available to the entire scientific community." (7). Therefore, open science could potentially be considered a subfactor under the category "collaboration", even though this is not explicitly mentioned.

While not appearing as an explicit high-level factor in an unpromoted search, there are a number of hits if open science terminology is explicitly included in Google search phrase. Many of these hits are from governmental agencies and policy makers: the UN refers to the call of three of its agencies, including the WHO, for more open science – this was undertaken in the context of the preparations of the UNESCO Open Science recommendation (10). The EOSC Secretariat also makes the connection between speed and open science, claiming that "[t]he ease of sharing means science is working at light speed and researches have already started working on a vaccine." (11) In a separate article the EOSC Secretariat presents the EOSC COVID-19 data portal which has been established to speed up vaccine development and allows more than 73 thousand users to the

platform, with 26,6 million requests being received since its launch on April 20, 2020. (20) Many of these documents mention in particular the early sharing of the COVID virus genome by Chinese scientists (10) (12) (16).

The OECD has also been active in this policy space, providing a collection of open science initiatives related to COVID-19 (14). The OECD has also issued a publication on "why open science is critical for combatting COVID-19" (15), which on the one hand stresses that open science can remove obstacles to the free flow of research data and ideas, claiming that "global sharing and collaboration of research data has reached unprecedented level" while on the other hand also outlines remaining challenges: "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." We will be coming back to these issues and proposed recommendations to mitigate them in the Google Scholar literature review below. Another OECD publication looks at the use of open data initiatives during the initial stage of the pandemic, finding that open governmental data (OGD) was important for communication efforts; however

> [d]espite its significant potential for crisis response, there is limited evidence that OGD initiatives drove concrete action beyond public communication efforts during the COVID-19 pandemic. Evidence indicates that many of these projects were data repositories or dashboards, with data analysis conducted most often in the form of data visualisations such as maps or charts.
>
> The results indicate that there has been a missed opportunity to use OGD to address the multi-dimensional implications of the COVID-19 pandemic with more sophisticated products or services during the initial stages of this lasting crisis. They also suggest that governments were not fully prepared or lacked the capacity to release relevant, high-quality datasets with the speed and quality necessary that can help address the crisis. Overall, the findings demonstrate that there are several policy challenges for governments to enable OGD re-use with high impact. (19)

As for Horizon Europe, an article in the official Horizon Magazine quotes a researcher saying that there is a lot more online sharing of data and sequences and a lot more tools available; with another source also highlighting the push towards open science, open data and pre-printing (16).

For the US perspective, one of the NIH institutes points to the fact that it encouraged its grantees to rapidly share COVID-19 research results and that, together with RDA guidelines and recommendations, this has led to an unprecedented volume of data being shared as well as additional practices that augment the immense potential of the data (17). However, they also point to some challenges such as low data sharing for some types of data.

2.2. Google Scholar Search Findings

The Google Scholar search query was run with the key words "open science covid", "open science covid speed", "open access covid", "open access covid speed", "open data covid", open data covid speed" "FAIR data covid", and "FAIR data covid speed", which, together with follow up research resulting in 111 documents.

Several of the documents define Open Science in the context of COVID; usually this categorisation includes open access and open data, sometimes also open source[6] [36] [101]; in most cases documents focus on either open access and open data, although some make reference to more broader open science enabling conditions.

Together with the increased speed of scientific production and the sheer amount of publications and data produced, open science components (open access /data sharing) are considered key factors of the scientific response to COVID-19. These factors are often considered co-dependent and are frequently but not always discussed together, either in editorials (which often voice a subjective view) or in empirical studies, sometimes through a "research on research" approach (even if it is not referred as such).

While some are unequivocally positive about open science changes [83] [92], others take a more ambiguous approach, and some are downright hostile. Jamali et al [46] for example note that open access, open science, and coronavirus are "megatrends with historical proportions" and state somewhat ambiguously:

> Speaking metaphorically, both Open Access and COVID-19 have invaded our lives without prior consent or permission… Of course, COVID-19 has affected the entire humanity, and Open Access has been more restricted in terms of having implications in the context of the academic community. But, we expect the ramifications of Open Access to also be long-term and far reaching, and change over time the rules of the game and expectations for publication in the academic community. In the same way that countries and human beings across the globe are now adapting to the realities of COVID-19, researchers and scholars will find ways to adapt to the realities and ramifications of Open Access. In fact, we hope and call on all scholars to strive in the realms of their specific countries and institutions to embrace Open Access given the advantages identified in this editorial relating to wider readership and circulation, and more efficient knowledge building and dissemination. [46]

In the following a more in-depth analysis of various aspects of these factors is provided.

2.3. Specific findings related to open access

a) bibliometrics: largely open, but…

A number of documents approach COVID-19 literature from the point of view of **bibliometrics.** COVID-19 related scientific publications include research articles, letters, editorials, notes, and reviews. As one example, amongst 21,542 documents in Scopus, 47.6% were research articles, 22.4% were letters, and the rest were reviews, editorials, notes and other. [88] Indeed, Elgendy et al [27] critically remark that there are too many opinion articles among COVID-19 medical literature. Bibliometric studies agree that **a high amount of COVID related literature is open** (e.g. [18], [19]), with the exact number varying according to the time the analysis was undertaken and which databases were used (Scopus, WoS, WHO etc). Aristovnik et al [2] note a divergence in open access rates between health sciences (82.9% in general and 82.0% for pure science) and life sciences (85.9% in general and 86.7% for pure science), while lower openness is

[6] Open source was not a specific search query though, and is therefore not discussed in depth here.

identified for physical sciences (67.9% in general and 50.0% for pure science) and social sciences and humanities (73.7% in general and 70.1% for pure science).

Torres-Salinas [95] also attempt to predict future developments in the COVID publication landscape and estimate continued growth of both open access and non-open access COVID-19 literature but with a higher growth rate of non-OA literature. Lee and Haupt [58] note a higher percentage of COVID-19 publications being open access as compared to non COVID literature:

> For each of the top 25 COVID-19 research-producing countries, there was a noticeably higher proportion of open-access articles on COVID-19 than during the past 5 years and on non-COVID-192020 publications during the same period, with all differences being significant except for one [58]

While the decision to provide COVID-19 literature in open access is seen as positive, **licencing** is sometimes criticised; a number of articles contain licences which make it clear that open access is potentially granted for a limited time only: e.g. as in the below:

> This article is made freely available for use in accordance with BMJ's website terms and conditions *for the duration of the covid-19 pandemic or until otherwise determined by BMJ*. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained. [13, own emphasis]

Even more significantly, Arrizabalga et al [3] found that 72.1% of all OA papers don't carry a specific license at all.

b) Volume of (open access) literature: infodemic

A large number of publications comment on the significant (rate of) increase in COVID-19 publications (e.g. [88], [94],[95]), with [102] referring to an estimate that 1 in 20 papers published in 2020 was related to COVID-19. A large number of authors discuss this in their contributions, with **most seeing it as a significant challenge to the scientific system**, and have named it an "infodemic" [6] or "paperdemic" [26], see also [12].

A number of articles in the Google Scholar sample describe the struggle of journal editors to deal with this deluge: "for example, at JAMA, from January 1 to June 1, 2020, more than 11 000 manuscripts were submitted, compared with approximately 4000 manuscripts submitted during the same period in 2019. Virtually the entire increase has been related to manuscripts focused on COVID-19" [5] Similarly,

> "PNAS experienced a surge of about 20% for the year, in fact, resulting in the publication of more than 3,600 research articles this year, compared with ~3,250 last year. Of the 30 most highly cited of the 3,600+ published papers, 90% were COVID-related. Most striking about the submissions was the diversity of disciplines they represented; researchers across the physical, biological, and social sciences pivoted to deploy their specific disciplinary skills to shed light on different dimensions of the pandemic." [7]

In this context Bauchner et al [5] point to a key question of debate, namely on **how to square the circle between scientific rigour and the need to speed up the publication process,** pointing to the (in their view) key role of editors in this process.

c) Key factors for speed: time to acceptance/publication and pre-prints

When **discussing the speed of and to publication there is a consensus in the literature that this has increased tremendously**. Two aspects are usually discussed in this connection: (i) attempts to increase speed in the traditional publication system and the (ii) "rise of pre-prints".

As concerns the former, Hurst's study of 669 articles concludes that

> medical journals have indeed drastically accelerated the publication process for COVID-19 related articles since the outbreak of the pandemic. Compared to articles published in the same journals before the pandemic, turnaround times have decreased on average by 49%" or 57 days on average, which is a statistically relevant difference. Some journals even show a decrease in publication time of over 80% compared to the pre-pandemic era. [41], see also [43].

Palayew et al also find a median time from receipt to acceptance of 6 days for COVID related journal articles [75]; similarly Helliwell et al find a *submission to acceptance* median of 5 days and a *acceptance to publication* median of 5 days for COVID related articles, which compares favourably with MERS literature [39]; Homolak et al [40] considers the decrement in submission to publication time to be 10 (similarly Kun [55]) and in some cases up to 15 times as compared to the pre-COVID state of play.

Palayew et al report a decreased acceptance time for non-COVID research, a finding which is contested by others [41]. In this context, several studies mention the measures taken by eLife to speed up publication.

Nevertheless, **a number of authors conclude that the traditional scientific system as such was not performing at the speed levels required to address the pandemic** [102]. The need to publish information at "pandemic speed" is given as the reason why many researchers turned to pre-prints [18]. Given that, particularly in the early stage of the pandemic, a significant percentage of the scientific literature were pre-prints ([56] claims nearly 40% but dropping to 26% by August 2020, see also [30,31],[88]) pre-prints are also seen by some as contributing to the information overload described above.

In fact, **articles that discuss the role of pre-prints are one of the most significant results returned for the search terms of the literature review:** they are acknowledged as a fast and open way to publish COVID information: Fraser et al [30,31] find that COVID-19 preprints are accessed more, cited more, and shared more on various online platforms than non-COVID-19 preprints. Homlack et al [40] positively note that

> preprint servers not only facilitated communication of scientific results, but also relieved part of the pressure from the journal editors and reviewers who had to expeditiously process all submitted articles, decide what to accept, organize rapid and high-quality peer-

> review, and make the data available to the rest of the scientists working on the problem worldwide.

However, the literature also contains a **significant number of critical voices** about this phenomenon (e.g. [68]). A systematic review of COVID-19 clinical studies concludes that COVID-19 clinical studies not only have a shorter time to publication but also lower methodological quality scores than control studies in the same journal [50]. Similarly, in their analysis of waste in COVID-19 research Glasziou et al [33] find that

> [m]any preprints are poorly reported, however. In systematically reviewing the proportion of asymptomatic covid-19 cases, we found the sample frame of most studies was unclear, missing cases were undocumented, and "asymptomatic" was undefined. We also identified disagreements between text and tables. Access to preprints has also led to irresponsible dissemination as flawed studies are picked up by the media.

Some therefore allege that the increased speed to publication in general and pre-prints in particular "risk the integrity of science in the era of COVID-19" [4] [18] [79]. While this is often presented as a viewpoint (in editorials) there are some articles which provide a more empirical analysis of the potential for misuse of pre-prints, including case studies. Koerber highlights a January 31, 2020 bioRxiv preprint publication that led to conspiracy theories by suggesting that SARS-CoV-2 originated in the laboratory through genetic engineering:

> [t]he team's findings were subject to rapid, intense criticism from the scientific community, with a formal rebuttal being published on March 8, 2020 and the pre-print was officially withdrawn, and it is noted as such on the bioRxiv website. While Open-science advocates tout this case as the best possible outcome of the rapid sharing of scientific information enabled by preprint servers such as bioRxiv because within 2 weeks, critics, however, have expressed an alternative perspective, emphasising the extent to which this preprint circulated and gained credence from audiences inclined to support a conspiracy theory during the short time it existed prior to being withdrawn. In fact, when this SARS-CoV-2/HIV preprint was published, its lead author, Pradhan, promoted it to more than 200,000 Twitter followers. [53]

Most concerns are directed on three interrelated aspects: peer review and retractions (as an indication of poor quality) and uptake in media reporting (e.g. [69]), social media [52] and policy [10]. In a blog, Heimstädt [37] discusses pre-prints under the heading "between fast science and fake news" and concludes that "speed and openness has also contributed to the ability of low-quality preprints to derail public debate and feed conspiracy theories." Somewhat stridently Oliveira and Jorge [26] argue that

> [t]he high-quality scientific journals were the last bastion to guide credible knowledge and strengthen decision-making. In other words, these high-quality journals were useful for distinguishing wheat from chaff. Nevertheless, with "publication fever", authors, editors and reviewers all lose the control of their function

Similarly, Bramstedt [13] even talks about a "carnage of substandard research" and Urban at et. al describe a situation in which

> [p]olicymakers and the media were left to sift through these preprints and arbitrate which scientific results to disseminate and promote rather than relying on the normally robust system of peer review. Erroneous and potentially dangerous research was posted to preprint servers and promulgated by the media, resulting in many preprint papers being retracted [102].

One often mentioned factor in these discussions is the **suspicion that the high volume to speed ratio compromises peer-review,** with some analysis even suspecting predatory publishing at work [86, also 18]. Chirico [17] holds the following concerns about "questionable" (his term) peer review practices:

> First, non-peer-reviewed pre-prints could be cited by peer-reviewed articles published in legitimate peer-reviewed journals, possibly leading to the spread of misinformation. Second, peer review quality may be compromised by rushing the process and assigning an excessive workload to peer reviewers, thus putting them under pressure and inducing psychological stress. Third, the overabundance of opinion papers and editorials hinders the discovery of valuable raw data and medical insight. Finally, some data, ideas, and content, preliminary and peer-reviewed, are constantly being disproved, outdated, and invalidated, which makes them candidates for corrections or retractions following post-publication peer review.

In his analysis, Horbach finds "notable diversity between COVID-19 and non-COVID-19-related articles, including fewer requests for additional experiments, more cooperative comments, and different suggestions to address too strong claims" [43].

For Jawaid and Jawaid [47] peer review simply is a time-consuming process for which there is no quick fix: "…if one starts compromising on peer review, then there will be no difference left in information carried by good quality peer reviewed medical journals and the information being made available through Social Media". Others have made concrete recommendations on how to improve peer-review, such as Da Silva et al [87]. Similarly, Urban et al call for the formation of rapid review boards which, in their view, would be able to square the circle between high quality peer review and speed [102, similarly also [45] [46]].

Others hold the view that pre-prints were never meant to replace peer review; rather, Vabret et al [104] describe how they implemented a **pre-print peer-review system**, where validated reviews are posted along the corresponding article on the pre-print server. Similar initiatives are also reported by [49] and [79]. Hocquet [38] is more radical and presents a grassroots web platform whose purpose is to potentially open a discussion forum about any scientific article-

Minari et al [68] consider that **the greater problem than pre-prints are peer review issues in high impact journals.** Like several other authors ([18] [87]) he takes retractions as an indicator, giving the example of the retraction of a paper in the Lancet in 2020, due to unreliable or non-existent data from Surgisphere [87]. Given the widespread references to this

retraction, this seems indeed to **have led some to question whether the current scientific system works well** or, as Mann puts it:

> How could 2 of the most prestigious medical journals with highly qualified editorial boards with countless years of editorial experience publish articles whose validity was uncertain at best and fictitious at worst. I believe that the retractions by the Lancet and the New England Journal of Medicine are symptomatic of subtle changes that have crept into the field of academic publishing over the past 60 years. [64]

Boschiero's analysis of retractions [11] concludes that "the COVID-19 pandemic…favoured the high level of retractions, including journals with the highest SJR and credibility in health science."

Minari et al [68] conclude that the failure by the current review system to spot the weaknesses of such articles can, at least in part, owe to the increased pressure on journals to quickly review and publish clinically relevant data on the pandemic. In this context Mann [64] talks about "warped speed" instead of "warp speed" and points to the fact that Retraction Watch had reported 20 COVID-19 papers as retracted, two as temporarily retracted and two papers containing misleading information by June 2020. Similarly, in their analysis Yeo-The and Tang find what they call an "alarming" retraction rate for scientific COVID-19 publications [108]. The topic even made it in the newspaper, with the New York Times publishing an article entitled "the Pandemic Claims new Victims: Prestigious Medical Journals" [80]:

> The hasty retractions, on the same day this month, have alarmed scientists worldwide who fear that the rush for research on the coronavirus has overwhelmed the peer review process and opened the door to fraud, threatening the credibility of respected medical journals just when they are needed most.

However, Abritis et al [1] caution that it is "premature to conclude that such [COVID-19] work is being retracted at higher rates than the rest of the literature."

2.4. Specific findings related to Open Data and Data sharing

Early on in the pandemic the link from open access to publications to data sharing was made, with Berenbaum [7] arguing that open access as the default also implies the need for open data.

While there are some that see a rush to share data in the pandemic [65] many comments in the early phase of the pandemic criticise in particular a lack of infrastructure to do so: as Cosgriff et al [22] put it in the Lancet: **"**Although health-care professionals use modern technology to communicate, never before has the failure to build robust data-sharing systems for large-scale near real-time analysis in health care been more obvious".

A number of researchers (see e.g. [62]) thus stress the need for data availability, with the COVID-19 Clinical Research Coalition Global coalition insisting that "funding agreements must mandate open collaboration and data sharing while protecting the rights of participants and patients… The FAIR guiding principles (Findability, Accessibility, Interoperability, and Reusability) for data should be implemented, and mechanisms put in place to enable equitable use and reuse of data [24].

Minari et al [68] are thus not alone in their view that "the term "openness" means much more than just open access to published data but must include all aspects of data generation, analysis, and dissemination along with other organisations and agencies than just research groups and publishers. Similarly Homolak et al note that

> [a]ccessible raw data would allow researchers all over the world to evaluate the statements being made and would thus represent the highest level of peer-review, ensuring the maximal level of information quality. As of the 28th of March, this kind of data is still not available to the large body of researchers switching focus to COVID-19 in order to provide help on this important global project….Regarding non-patient-related data on COVID-19, organization and availability are also still suboptimal—nonetheless, some improvements have been made. [40]

Analysing the data underlying COVID publications, Lucas-Dominguez et al [60] find that only 13,6% of articles share underlying data; furthermore

> [t]he analysis of supplementary material showed that three-quarters of the documents were PDF and DOC, containing mostly textual or graphic materials complementary to the research, and a percentage that barely reached 10% (73 papers) were files with reusable data formats (xls and csv), which is equivalent to 1.2% of the 5,905 records published and analysed in this work.

They thus conclude that data sharing is not a common practice, even in health emergencies, such as the present one. Sixto-Costoya et al come to similar conclusions in their analysis of data sharing in emergency medicine [96]:

> [i]t has been observed that 19 out of the 24 journals contained in the emergency medicine category of Journal Citation Reports are also located in PubMed Central (PMC), yielding a total of 5983 articles. Out of these, only 9.4% of the articles contain supplemental material.

Pecoraro analyses the FAIRness of COVID data and also finds

> differences across datasets…in their accessibility and findability….considering the interoperability and reusability principles, all datasets lack the use of a data model as well as of standards for the representation of data and metadata. Moreover, the absence of a clear data flow that describes the provenance of data makes it difficult to integrate data and perform a multi-country analysis. Positively, data are open and may be reused for statistical purposes without requiring authentication to relevant websites. [77]

Balkanyi et al are more pronounced in their assessment that "**the quality of data is not always good and rarely FAIR"** [9]. Researchers working in different but related areas voice similar concerns. In a piece entitled "COVID-19 and the research scholarship ecosystem: help!" Mother holds that data sharing practices remain elusive in biomedicine:

> What is concerning, and insufficiently discussed in the literature, is the lack of direct access to data underpinning much of the COVID-19 research output, in general, and the all-important results of the vaccine trials, in particular….None of the data underlying the vaccine trials is directly and easily available to the scientific community, patients or the broader community. Lack of access to the trial data might explain why none of the vaccine trials were published in BMJ or PLOS Medicine. Both journals have the highest data sharing standards of journals, to date [70]

He therefore concludes that **while there have been valiant efforts to ensure easy access to COVID research and sharing of its underlying data – the Wellcome initiative early on in the pandemic endorsed by hundreds of organizations – this has not resulted in meaningful change**. [70]

Besancon et al [9] refers to a recent review of registered COVID19 trials which showed that only 16% of 924 included trials indicated a willingness to share the trial data but also refer to positive examples such as the MIMIC database. Similarly, Yamada et al describe a centralised repository of individual-level information on patients with laboratory-confirmed COVID-19 in China [107]. Furthermore, Treweek et al point to the fact that between 23 January 2020 and 8 March 2020 alone, there were 382 new registered COVID-19 trials on the World Health Organisation International Clinical Trials Registry Platform (ICTRP) and by 31 March 2020 the ICTRP COVID-19 trial list had grown to 586. [99] In this context, the journal "Trials" announced its intention to make publishing protocols for COVID-19 trials that have ethical approval much fast and much simpler [99]. Besancon et al [9] provide a number of recommendations for data-sharing policies from a journal perspective.

Two publications have specifically focused on **mental health data** [25],[73] **with very similar findings**. Demkowicz et al [25] hold that robust and open methods may have been compromised by speedy responses and Nieto et al's systematic review [73] finds low compliance with open science recommendations. More positively, Yamada et al [108] describe a dataset on the psychological and behavioural consequences of the pandemic.

Branching out into data related subjects, a **number of publications address AI** [10, 85, 93, 97] or similar areas such as deep learning [112] and machine learning [78]. A number point to problems with the available data and how it may lead to bias. Röösli et al [85] recount that

> the most frequent problems encountered were unrepresentative data samples, high likelihood of model overfitting, and imprecise reporting of study populations and intended model use. There is an urgent need to balance the rapid dissemination of evidence to guide clinical decision-making with unbiased, high-quality models that truly benefit all populations in the COVID-19 era.

Shuja et al [93] see a need for open-source curation and extraction in multiple directions and note "that the main challenge towards data-driven AI is the opaqueness of data and research methods." Similarly, Sonntag sees the need for "a systematic collection of patient information and medical literature in a digital format, and machine learning solutions for end users." [97] However, Pfeiifer-Smadja et al [78] point out that open sharing of clinical databases requires significant

care to properly manage regulatory and data privacy issues. Otherwise, machine learning may not keep its promises to help fight the virus. This is why **open intellectual property licences**, such as the open covid pledge are important [see 20].

Pfeiifer-Smadja et al also criticise that the lack of large healthcare, clinical, imaging, and genetic public repositories leads each institution to locally develop its own analytical pipeline on its own small dataset, which significantly limits the generalizability of the results [78]. Such concerns over a fragmented landscape of data are also reflected in Ewers et al's contribution on clinical trial data. [29]. They note that before the advent of the COVID-19 crisis, there was no sense of urgency to share and point to potential privacy barriers, also flagged up by McMahon et al [68]. However, they note that

> several anonymization procedures have been proposed, and metrics for the quantitative assessment of the risk of reidentification have been developed. Although entirely irreversible anonymization of data remains difficult, a balance between the benefit of sharing the data and risk of personal harm by disclosure must be achieved. [29]

As concerns an overall estimation of the success of data sharing Homolka [40] states:

> Preliminary data suggest that COVID-19 pandemic resulted in generation of a large amount of scientific data, and demonstrates potential problems regarding the information velocity, availability, and scientific collaboration in the early stages of the pandemic… In conclusion, we believe the scientific community could have used the data more efficiently in order to create proper foundations for finding new solutions for the COVID-19 pandemic. Moreover, we believe we can learn from this on the go and adopt open science principles and a more mindful approach to COVID-19-related data to accelerate the discovery of more efficient solutions

Even in 2021 Gardner et al [32] still see the need for open public data standards and sharing in light of what they call the "disjointed" public health response to the pandemic. They hold that "without such a system in place, our ability to respond effectively to the inevitable future pandemics will be limited, and the cost will be human lives".

2.5. Resources, tools, and national case studies

A number of documents also provide examples of open COVID resources (literature / datasets), e.g. [74] and/or catalogues of such resources [2], infrastructures [71] and tools [100], developed in different countries [8] [14] [21] [44]; country related studies include Turkey [63] and Taiwan. [34] provides a collection of government responses, the data of which are available in the open.

2.6. Framework conditions and other open science components

Several documents refer to open science practices in the **framework of science for public good or as public good** ([15] [16] [54] [106]), sometimes with anti-capitalist leanings. Early in the pandemic the LA times reported that "that COVID-19 could kill the for-profit science publishing model – that would be a good thing" [42]. Krishna [54] holds that

> the whole paradigm of open science and its social contract is being challenged by various "enemies" or adversaries such as (a) market-based privatized commercial science, (b) industry 4.0 advanced technologies, and (c) a "new iron curtain" on the free flow of science data and information… A major threat to open science has come from what has come to be known as platform capitalism which refers to the activities of companies such as Google, Facebook, Apple, Microsoft, Uber, and Airbnb operating as platforms. They are monopolizing data of all sorts to use in their commercial and corporate affairs and trading.

Lee and Haupt look at open science and COVID-19 through the lense of scientific globalism, concluding that

> scientific globalism occurs differently when comparing COVID-19 publications with non-COVID-19 publications during as well as before the pandemic. Despite the tense geopolitical climate, countries increased their proportion of international collaboration and open-access publications during the pandemic. However, not all countries engaged more globally. Countries that have been more impacted by the crisis and those with relatively lower GDPs tended to participate more in scientific globalism than their counterparts. [58]

The same analyses also showed that the higher the GDP, the lower the odds of international collaboration and open-access publishing during a global crisis.

A number of authors also see the **need to go beyond data sharing and open access to publications and embrace other components of open science,** with Joseph [48] identifying **research assessment in particular**. Kayvan et al [52] concur as follows:

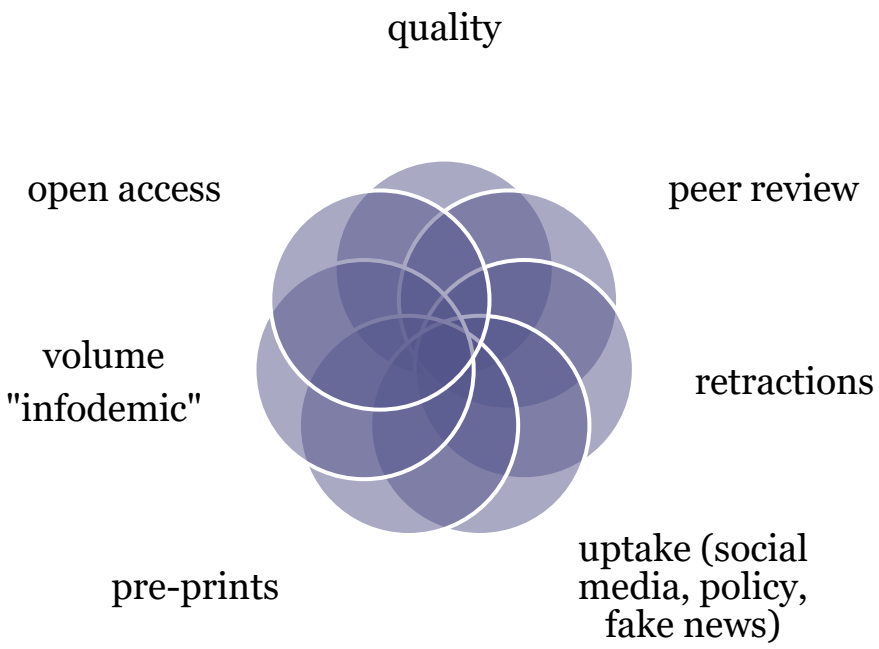
> Changes in incentive structures and forms of recognition are necessary, from taking preprint and other forms of research into account, to crediting data analysts, postdoctoral researchers, and other under-recognised contributors whose work is essential to the research enterprise. Open data sharing and collaborations between groups show a potential way to help eliminate the waste, burdensome bureaucracy, and duplicated effort that have been seen in COVID-19 research.

Through interviews with South Korean and Australian experts Shmagun et al [91] highlight factors that either enable or limit the impact of Open Science during a public health emergency, such as the COVID-19 outbreak. Institutional and regulatory factors are perceived as the most important ones by the interviewees. They note that "even the developed countries, which had already had pre-pandemic commitment to OS development, have shown the lack of comprehensive and consistent OS policies, inadequacy of cross-sectoral coordination mechanisms, and insufficiency of current incentive structures for researchers to pursue greater openness and collaboration" and that "many initiatives launched during the current pandemic, such as open access to coronavirus-related publications, seem to be a temporary response to the crisis rather than the start of more sustainable structural changes in research culture."

2.7. Conclusions

Looking at the documents identified by the regular Google search and comparing them with the literature review from the Google Scholar search we see **that policy makers and related organisations generally provide a positive assessment of open science practices in COVID** with the EOSC secretariat claiming that "the ease of sharing means science is working at light speed" (11), though some also outlined some challenges (most notably the OECD). **The academic literature, by contrast, is more nuanced and sometimes more sceptical.** Many of the practices adopted during the pandemic are seen as important, including the need for speed, but at the same time the review reveals **a number of concerns about whether quality has been sacrificed for speed**. Several aspects that have come to the fore with the need to speed up research, such as expedited peer review and the rise of pre-prints have been met with scepticism by a number of (but not all) authors, perhaps also reflecting a certain amount of orthodoxy among a profession which should be open to trying out new approaches.[7] In summary the figure below shows the issues often mentioned together with the speed of COVID research and open science

Figure: factors mentioned in relation to the speed of COVID research and open science



For Demkowicz et al [25] open science practices have clearly not been abandoned, but it is possible that desire for more knowledge quickly meant open science was not applied with the same rigor [25]. Open Science advocates argue that these challenges are, in fact, not the result of too much

[7] A number of these issues were also discussed in an event entitled "what has the pandemic taught us about 'open science' held as part of the Berlin Science Week 2021. See https://berlinscienceweek.com/event/what-has-the-pandemic-taught-us-about-open-science/

open science practices but are the consequences of open science practices being only partially or sub optimally embraced [9].

In any case, it is clear from the literature review that while COVID-19 has brought lots of changes, **we are not yet living in an Open Science or an Open Data world**. Rios et al [84] claim that

> the scientific community is still currently debating the pros and cons of full implementation of complete data sharing, despite the existence of an overall belief that open science is the natural step to take for scientific progress. This is due to a discrepancy born out of 'what should be done' (ideally) and 'what is' or 'what will actually be done' (realistically). So, to understand this divergence, we should have in mind, what are the concrete advantages and disadvantages from both the point of view of the individual scientist and from scientific progress.

Looking into the future some authors voice concerns as to whether science after COVID will indeed be "faster, better, stronger" [90]. To the author of this review the perspective taken by the Lancet [57], which sees **old and new ways of doing science as co-existing for some time** is perhaps the most realistic one. However, this also means that tensions between progressive (sometimes radical) members of the scientific community, who consider shortcomings to be the fault of too little and not too much open science, and conservative researchers, who would like to retain or expand the role of traditional gatekeeps as a bastion for quality, will continue.

3. Survey Results

Based on the outcomes of the literature review the survey was designed and after review and input by the client implemented in Survey Monkey. The initial target group were publicly funded COVID research projects. However, given that the response rate was extremely low for this target group (23 responses), a second target group was added, namely those authors that have published COVID related papers with Frontiers. This resulted in an additional 5,790 email addresses to which the survey was sent. The overall number of responses was 155 as of July 20. After a reminder was sent, the overall response rate rose to 208 by August 8, 2022. 34% of the respondents are from the medical sciences, 30% from social sciences and humanities and 29% from public health (7% other).

3.1. General factors and policies

From the high-level factors identified in the literature review as playing a role in explaining the unparallel speed of COVID research, the factor with by far the most survey responses (78% or 160 responses) is the worldwide collaboration between all actors. As discussed in the literature review this could be considered to *implicitly* include data sharing. This is followed by fast-track funding through dedicated instruments from research funders (51% or 105 responses). Access to previous research on SARS and MERS comes in third place (37,56% or 77 responses).

Q3: Which of the following factors do you consider important for the unparalleled speed of COVID research? (check all that apply)

Answered: 205 Skipped: 3

ANSWER CHOICES	RESPONSES	
Access to previous research on SARS and MERS	37.56%	77
Fast track funding through dedicated instruments from research funders around the globe	51.22%	105
Worldwide collaboration between all actors (researchers, manufacturers, distributers, authorities and a range of public and private institutions)	78.05%	160
New Technology (mRNA but also viral vectors)	23.41%	48
Speeding up of processes in regulation (as regards clinical trials and regulatory approval) and manufacturing (parallel phasing)	36.10%	74
Sharing of the COVID genome	29.27%	60
Other (please specify)	7.32%	15
TOTAL		539

Public documents from funders and policy makers were on occasion somewhat self-congratulatory in the self-assessment of their COVID activities. When asked how effective researchers consider these mandates to be, the average rating was slightly more than 3 out of 5 stars (weighted average 3,24).

3.2. Access to scientific publications and related issues

When specifically asked about the importance of access to scientific publications as a contributing factor to the unparalleled speed of COVID research (1 star not at all, 5 stars very much so) researchers provided a weighted average of 4,19 stars, with 80% of respondents providing a 4 or a 5 star rating.

Q5: In your research, how important was access to scientific publications as a contributing factor to the unparalleled speed of COVID research? (1 star not at all, 5 star very much so)

	1	2	3	4	5	TOTAL	WEIGHTED AVERAGE
☆	0.48% 1	6.28% 13	13.53% 28	32.85% 68	46.86% 97	207	4.19

Answered: 207 Skipped: 1

In the survey, the same percent of respondents said they had more access to COVID literature than to non-COVID literature and that the levels of access were the same (47% each). Respondents said that on average 72% of the literature sources they used were open. 82% consider themselves very satisfied or satisfied with this level of access. This confirms that findings of the literature review that a high amount of COVID related literature was indeed open. In this respect, then, the mandates and policies of funders and policy makers can be judged to have been effective.[8]

Q6: In your estimation, did you have more, the same or less access to scientific literature in your COVID research compared to non-COVID research?

Answered: 208 Skipped: 0

ANSWER CHOICES	RESPONSES	
More	47.12%	98
About the same	47.60%	99
Less	5.29%	11
TOTAL		208

[8] For open access to publication, criticism focuses more on the ability of publishers to shut down access after the pandemic, due to the nature of the licencing agreements (see literature review).

Q8: In general, how satisfied have you been with access to scientific publications during COVID? (Scale 1: not at all satisfied 5 very satisfied)

Answered: 207 Skipped: 1

ANSWER CHOICES	RESPONSES	
Very satisfied	27.54%	57
Satisfied	54.59%	113
Neither satisfied nor dissatisfied	14.49%	30
Dissatisfied	2.42%	5
Very dissatisfied	0.97%	2
TOTAL		207

Concerning peer review – in particular measures taken to speed it up – this was a source of contention and debate in the literature review. In the survey, results related to peer review tended to be on the more positive side:

- 62% considered the effect of the increased speed of COVID publications on peer review as positive or very positive.
- 73% strongly agreed or agreed that light of their experience with COVID research, there is a need to reform peer review.

Q9: What do you consider to be the effect of the increased speed of COVID publications on peer review?

Answered: 206 Skipped: 2

ANSWER CHOICES	RESPONSES	
None of the above	1.94%	4
Very positive effect	24.27%	50
Rather positive effect	37.86%	78
Neutral	21.84%	45
Rather negative	10.68%	22
Very negative	3.40%	7
TOTAL		206

Q10: In light of your experience with your COVID research, do you agree that there is a need to reform peer review?

Answered: 207 Skipped: 1

ANSWER CHOICES	RESPONSES	
Strongly agree	31.40%	65
Agree	41.55%	86
Neither agree nor disagree	21.74%	45
Disagree	4.83%	10
Strongly disagree	0.48%	1
TOTAL		207

Mixed attitudes emerged concerning the question on whether speeding up speeding up of COVID research has had a negative effect on the quality of research, a key matter of debate in the literature review: 40% strongly agreed or agreed with this statement but 35% are undecided ("neither agree or disagree") with 25% disagreeing strongly or disagreeing.

Q11: Do you agree that the speeding up of COVID research has had a negative effect on the quality of research (e.g. retraction rates)?

Answered: 206 Skipped: 2

ANSWER CHOICES	RESPONSES	
Strongly agree	8.74%	18
Agree	31.07%	64
Neither agree nor disagree	35.44%	73
Disagree	19.90%	41
Strongly disagree	4.85%	10
TOTAL		206

3.3. Access to scientific data and related issues

How important was access to scientific data as a contributing factor to the unparalleled speed of COVID research? On a scale from 1 to 5, the weighted average from the respondents is 3,88, which can be interpreted as an important factor but hardly the panacea it has been considered in some statements and documents, although 68% of respondents did provide a 4- or 5-star rating.

Answered: 203 Skipped: 5

	1	2	3	4	5	TOTAL	WEIGHTED AVERAGE
☆	2.46% 5	7.39% 15	22.17% 45	35.47% 72	32.51% 66	203	3.88

Q12: In your research, how important is access to scientific data as a contributing factor to the unparalleled speed of COVID research? (1 star not at all, 5 star very much so)

64% of participants were very satisfied or satisfied with the level of data access, which is significant but also somewhat lower than the satisfaction level with access to publications. Although the literature review seemed to indicate otherwise, a majority of respondents were very satisfied or satisfied with the level of access specifically to data underlying published articles (59%). A slightly lower number – but still a majority (53%) were very satisfied or satisfied with the quality of the available COVID data.

Q13: In general, how satisfied have you been with access to COVID data during the pandemic? (Scale 1: not at all satisfied 5 very satisfied)

Answered: 205 Skipped: 3

ANSWER CHOICES	RESPONSES	
Very satisfied	18.05%	37
Satisfied	46.34%	95
Neither satisfied nor dissatisfied	29.27%	60
Dissatisfied	4.39%	9
Very dissatisfied	1.95%	4
TOTAL		205

Q14: How satisfied have you been with sharing of COVID data underlying published articles? (Scale 1: not at all satisfied 5 very satisfied)

Answered: 205 Skipped: 3

ANSWER CHOICES	RESPONSES	
Very satisfied	8.78%	18
Satisfied	50.24%	103
Neither satisfied nor dissatisfied	36.59%	75
Dissatisfied	3.41%	7
Very dissatisfied	0.98%	2
TOTAL		205

Q15: How satisfied have you been with the quality of the available COVID data? (Scale 1: not at all satisfied 5 very satisfied)

Answered: 204 Skipped: 4

ANSWER CHOICES	RESPONSES	
Very satisfied	6.37%	13
Satisfied	46.57%	95
Neither satisfied nor dissatisfied	37.25%	76
Dissatisfied	9.31%	19
Very dissatisfied	0.49%	1
TOTAL		204

It is notable, however, that on average the respondents considered 58% of datasets they used in their COVID research to be open access, a significantly lower number than open access publications used in COVID research. 43% said that more COVID datasets than non COVID datasets were open, and 51% considered the amount the same. In total, 67% strongly agree or agree that the COVID pandemic has improved data sharing practices. However, a large majority (90%) also sees the need to further improve data sharing practices.

Q17: In your estimation, did you have more, the same or less access to scientific datasets in your COVID research compared to non-COVID research?

Answered: 202 Skipped: 6

ANSWER CHOICES	RESPONSES	
More	43.07%	87
About the same	50.99%	103
Less	5.94%	12
TOTAL		202

Q18: Do you agree that the COVID pandemic has improved data sharing practices?

Answered: 205 Skipped: 3

ANSWER CHOICES	RESPONSES	
Strongly agree	20.98%	43
Agree	46.34%	95
Neither agree nor disagree	24.88%	51
Disagree	6.34%	13
Strongly disagree	1.46%	3
TOTAL		205

Q19: Do you agree that further improvements in data sharing practices are still needed?

Answered: 205 Skipped: 3

ANSWER CHOICES	RESPONSES	
Strongly agree	36.59%	75
Agree	53.17%	109
Neither agree nor disagree	9.27%	19
Disagree	0.98%	2
Strongly disagree	0%	0
TOTAL		205

3.4. Other open science practices

On the specific issue of licencing[9], open IP licences were considered by 72% to be very important or important. More generally, a "super majority" of 93% strongly agrees or agrees that scientific information (publications/data) produced by public funding are a common good and should therefore be available as open as possible. Furthermore, 71% strongly agree or agree that open science practices beyond open access to publications and data sharing are necessary. Specifically, 79% strongly agree or agree that changing the way we assess researchers and research projects is important. Nevertheless, when asked whether the scientific system has worked well addressing the COVID pandemic (Scale 1: not at all, 5 very much so) the weighted average from the respondents is 3,63%, a rather positive assessment. The impression that the survey sample was open to experimenting with new approaches is confirmed by the last question. 68% agree with the statement that we need more open practices to ensure the quality of research and only 18% agree with the statement that we need traditional gatekeepers to ensure the quality of research.

[9] The type of licence provided with datasets can be key as regards the re-usability of such data (i.e. the R in FAIR).

Q20: How important for the speed of results do you consider open intellectual property licences (for publications and/or data)?

Answered: 202 Skipped: 6

ANSWER CHOICES	RESPONSES	
A great deal	27.23%	55
A lot	45.05%	91
A moderate amount	22.28%	45
A little	2.48%	5
Not at all	2.97%	6
TOTAL		202

Q21: Do you agree that scientific information (publications/data) produced by public funding are a common good and should therefore be available as open as possible?

Answered: 207 Skipped: 1

ANSWER CHOICES	RESPONSES	
Strongly agree	51.21%	106
Agree	42.03%	87
Neither agree nor disagree	4.83%	10
Disagree	1.93%	4
Strongly disagree	0%	0
TOTAL		207

Q22: Do you agree that open science practices beyond open access to publications and data sharing are necessary?

Answered: 205 Skipped: 3

ANSWER CHOICES	RESPONSES	
Strongly agree	28.78%	59
Agree	42.44%	87
Neither agree nor disagree	23.90%	49
Disagree	4.39%	9
Strongly disagree	0.49%	1
TOTAL		205

Q24: In light of your experience with COVID research, do you agree that changing the way we assess researchers and research projects is important?

Answered: 206 Skipped: 2

ANSWER CHOICES	RESPONSES	
Strongly agree	28.64%	59
Agree	50.0%	103
Neither agree nor disagree	16.99%	35
Disagree	3.88%	8
Strongly disagree	0.49%	1
TOTAL		206

Q25: All in all, do you consider that all in all the scientific system has worked well in addressing the COVID pandemic? (Scale 1: not at all, 5 very much so)

Answered: 203 Skipped: 5

	1	2	3	4	5	TOTAL	WEIGHTED AVERAGE
★	3.45% 7	8.87% 18	26.60% 54	43.35% 88	17.73% 36	203	3.63

Q26: Which statement do you agree with more:

Answered: 204 Skipped: 4

ANSWER CHOICES	RESPONSES	
None of the above	14.22%	29
We need more open practices to ensure quality of research	67.65%	138
We need traditional gatekeepers to ensure the quality of research	18.14%	37
TOTAL		204

3.5. Conclusions

The survey shows that a large majority of the respondents was satisfied with access to COVID relevant literature and in this regard the activities by funders and policy makers can be considered a success.[10]

The measures taken to speed up peer review were controversially discussed in the literature; in the survey speeding up peer review of COVID related literature was seen in a positive light and the survey participants saw a need to continue reforming peer review. Respondents were rather split on the question whether speeding up the production of knowledge had negative effects on the quality of research, another issue hotly debated in the literature.

As concerns access to COVID data respondents were by and large also satisfied, although the percentage of those very satisfied and satisfied is significantly lower than for access to publications. While two thirds of respondents agreed that the COVID pandemic has improved data sharing practices, an even larger percentage (91%) see the need to further improve data sharing practices.

We also found that a super majority of respondents (93%) are generally in favour of scientific information (publications/data) being available in open access. Furthermore, 69% agree that further open science practices beyond publications and data are needed, with 77% agreeing that research assessment needs be changed. Overall, however, the respondents were positive about the way the scientific system has coped with the challenges of COVID.

Looking again at the key question of the study, that is the importance of open science as a contributing factor to the unparalleled speed of COVID research we see an interesting divergence in the evaluation of open access to publications and open access to data. Respondents ranked the importance of open access to publications in this regard much higher than open access to research data:

[10] although this could be undermined by publishers removing open access in the post pandemic era.

Table: importance of open science as a contributing factor to the unparalleled speed of COVID research

Open Science element	weighted average	Respondents allocating 4 or 5 stars
OA publications	4,19	80%
Open data	3,88	68%

Scale: 1-5 stars (1-not important, 5-very important)

This issue was further explored in the interview phase of the project.

Generally, the results of the survey provide an important corrective to the literature review, in which many authors were more sceptical about the successes of open science as related to COVID. By comparison, the survey sample showed a more open attitude, with a majority of 68% agreeing that we need more open practices to ensure the quality of research and only 18% agreeing with the statement that we need traditional gatekeepers to ensure the quality of research. Since nearly all respondents were drawn from Frontiers authors, this may indicate that this group in itself has a more open science attitude.

4. Interview results

Based on the results and learnings from phase 1, in phase 2 qualitative interviews with six stakeholders on their experience with open science practices in the COVID context were conducted. These stakeholders – scientists active in COVID research – were selected based on a list of authors, which published on COVID with Frontiers. From this data, a long list of 10 potential interviewees with complementary experience and global distribution as well a gender balance was established, with a further backup list of 8 interviewees. The first 10 potential interviewees were contacted multiple times, resulting in a response rate of 50%. An additional interviewee was therefore recruited from the backup list. The following people were therefore interviewed in the period between End of October 2022 and early February 2023:

- **Interview 1:** male, United States Associate Professor, Neuroscience & Cell
- **Interview 2:** male, Indonesia Universitas Gadjah Mada, Genomic Surveillance of SARS-CoV-2
- **Interview 3:** male, Germany Director General, European research infrastructure
- **Interview 4:** female, United States, Independent Nurse Researcher
- **Interview 5** male, Serbian Associate Research Professor (Senior Research Associate) Immunology and Immunoparasitology, COVID project on cellular immune response
- **Interview 6** male, Norway Professor Faculty of Health and Social Sciences, responsible for microbiology, epidemiology, and infectious disease control

Based on the results of phase 1 (see above) an interview guide was developed with the aim of obtaining feedback on the preliminary results as well as to deepen key aspects. Interviewees were provided with a project Information and Informed Consent Form, with the twin aim of providing first results to the interviewees and ensuring that the interview and the resulting data adheres to the General Data Protection legislation. In the following we provide a summary of the input per each interview question for each interviewee.

4.1. Feedback on the main factors identified in the literature for the speed of COVID research

Interviewee 1 pointed out that the approval of the COVID-19 vaccine was influenced by political factors as well as scientific research. The regulations were made more flexible to approve the vaccine quickly, and this was a political agreement more than a scientific one. The urgency and fear of the spread of COVID-19 also played a role in the approval process. The political will was there because of the necessity to address the unknown potential consequences of the pandemic.

For interviewee 2 the most important factor was simply that COVID turned out to be a pandemic, i.e. a new disease that has affected almost every country in the world, making it a global concern not just for health workers but for scientists, researchers, and ordinary people as well. Further factors were the lack of available treatments and vaccines, and the urgency to find solutions, which led to a surge in research efforts. Genomic surveillance also played a crucial role in understanding the virus and developing effective solutions.

Interviewee 3 did not notice a significant increase in worldwide collaborations for COVID research; he prioritised quick access to new data on potential therapeutics under preclinical or discovery phases from academic institutes they are already collaborating with.

Interviewee 4 stressed the importance of worldwide collaboration (in her case 50 researchers) from various fields. Despite the chaotic and conflicting data that emerged during the pandemic, the world's greatest minds sought out collaboration to make sense of the information. The speaker also notes that (in her view) poor quality articles and political considerations influenced what was published during the pandemic.

The most important factor for interviewee 5 was his desire to help during the COVID-19 pandemic. His team monitored the literature in a think-tank like structure. They note the importance of open sources but also express concern about the potential for misleading information.

Interviewee 6 confirmed the importance of international collaboration in response to the pandemic. Existing networks allowed them to connect with colleagues outside of Norway, leading to sharing experiences, publishing, and establishing research groups. He stresses that such collaborations were essential for gaining a global perspective on the pandemic.

4.2. Feedback on access to scientific publications and speed of COVID research

Interviewee 1 expressed agreement with the benefits of increased access to science and faster publication times, but also suggests that there may have been a decrease in the rigor and reproducibility of data as a result. He notes that some papers have been published with poor quality and that some had only one case, which may not provide a good representation. He argues that when dealing with topics that involve sick patients and politicians, it is important to consider the impact of the research on people's fears and concerns. He suggests that it is important to balance the positive impacts of speed with maintaining scientific standards.

Interviewee 1 also provided two examples of limitations to accessing scientific literature. The first example is a colleague in Pakistan who did not have access to certain journals related to tuberculosis, which is a major issue in this country. The second example are clinicians in developed countries who work in hospitals that are not affiliated with universities and do not have the same access to journals. While he personally had the access he needed in the United States, he acknowledges that limitations exist in other countries and contexts.

Interviewee 2 noted that it is fortunate that COVID-19 publications, even pre-prints, are open and accessible to everyone, which is important for the development of knowledge around the disease. Journals are requiring preprints while undergoing peer review, and the interviewee values being able to access publications about COVID-19 freely. Additionally, the interviewee mentions that after publishing a paper about the self-genomic of SARS-CoV -2 in their province, a potential collaborator reached out to them for future research, potentially due to the accessibility of the publication.

Interviewee 3 did not have difficulty accessing COVID-related research publications as they were mostly open access. He was surprised by how fast things were being published: they were involved in one publication where a partner tested some of their compounds and the data was made publicly available within two to three months. He did not encounter any issues with collaborators not having access to the research.

Interviewee 4 considered Open Access publishing “a mixed bag” as it is not always free for authors. Most Open Access publications charge authors a fee, which can be a significant hurdle for independent researchers. The ability to access articles also varies, with not all desired articles being free, but with some researchers having access through subscriptions to various sources.

Interviewee 5 did not face any access problems in Serbia, simply because his institutions only have open access to begin with and rarely have the money to pay for access. The team struggled to pay for the open access publication of their own COVID research and ended up paying from their own money. The interviewee considers the quick dissemination of COVID research as very important and worth the cost.

During the pandemic, interviewee 6 did not have any difficulty accessing literature as he lived in a country with good access to most literature. However, he did have colleagues who had difficulty accessing literature and he shared necessary literature with them. The interviewee found that

most of the important literature was open access, and he used a service called BioRICS which provides access to most of the literature he worked with. He believes that most of the peer-reviewed papers were open access, and many big publishers have policies of open access to all COVID-related research.

4.3. Feedback on quality concerns, pre-prints and peer review

Interviewee 1 expressed concern about pre-prints and journals that allow papers to be published without review. He believes that releasing information without review to a scared public is not smart and could lead to misinformation. The interviewee also believes that even if papers are retracted, the information can still lead to wrong information being perpetuated in the media. He suggests that this is a complicated problem that needs to be addressed by the scientific community, and he encourages more rigorous review processes to avoid false or misleading information being disseminated.

Interviewee 2 is in favour of pre-prints because they allow for quick dissemination of research and provide valuable knowledge for researchers, scientists, and stakeholders. The World Health Organization (WHO) uses pre-prints as references for their weekly updates on COVID-19. The interviewee notes that while pre-prints are not final publications, they provide prior knowledge that is important when quick action is needed to tackle the virus. The interviewee gives an example of how pre-prints have helped understand how Omicron and Delta affect COVID severity, spreading, and antibodies. Waiting for final versions of journals is not practical in urgent situations like these.

Interviewee 2 believes that peer review has already been reformed to some extent, with journals offering open, blind, and transparent peer review. He thinks that transparent peer review is currently the most appropriate because it allows reviewers to choose whether to publish their name in the review. Some may prefer not to be named due to negative comments or competition.

Interviewee 3 remarked that a lack of proper review can of course lead to low-quality data, but this is not unique to COVID-related research. Even peer-reviewed papers can vary in quality depending on the reviewer. He believes it is helpful for publications to note that they were fast-tracked so that readers can interpret the data critically. Interviewee 3 stated that the peer review system works in principle, but there are practical problems with getting reviewers to do the work. He mentions similar issues in finding reviewers for research proposals and is considering paying them for their efforts. He believes that the main problem is that people are too busy or occupied with other tasks.

Interviewee 3 was positive about pre-prints, as long as it is clear to the reader that the data has not been peer-reviewed and that they must be interpreted it at their own risk. He believes that it is better to have pre-prints published as quickly as possible, but ultimately it is important to have peer-reviewed publications.

Interviewee 4 considers pre-print servers a good thing, as they allow for open peer review with a variety of experts, which can improve the quality of work before it is set in stone. The interviewee

considers pre-prints as more reliable than work that has gone through peer-review and been published, as editors of journals can have agendas. ResearchGate is not officially a pre-print server, but people often put their work on the site and get comments that can improve the quality of their work.

Interviewee 4 had content changes by editors made after the peer review process. In her opinion, the peer review system is therefore broken, since the final say should be with peer reviewers. The interviewee suggests that an open review process can be an advantage of doing a pre-print. The interviewee also notes that many journals are desperate for peer reviewers and allow authors to suggest reviewers, which can create a conflict of interest. The interviewee suggests a blinded peer review system where nobody knows who is doing what for a true peer review system.

Interviewee 5 considered pre-prints to be important in scientific research but also highlights the need to be cautious in interpreting the information provided in pre-prints, as they are not peer-reviewed. The author emphasises the importance of monitoring the groups publishing pre-prints and their publication history to ensure the reliability of the information (as they did in their own work). The interviewee prefers peer-reviewed research but acknowledges the value of monitoring preprints in certain instances.

Interviewee 5 suggested that that it would be better if peer reviewers were paid to speed up the review process. The interviewee acknowledges that although the peer review process is necessary for critical input, it is not perfect and can be improved with multiple reviews. They suggest that the involvement of open AI could help make the process faster, but the peer review system is the best we have for now.

Interviewee 6 mentioned that peer review is not always indicative of quality during the pandemic and that he relied on his own expertise to judge the scientific quality of papers. In this sense he considered all papers as pre-prints and handles them with care. The interviewee likes open peer review and believes it provides a quality indicator. He also expressed dissatisfaction with the fact that journals make a lot of money from the free work done by scientists. The interviewee thinks that the peer review system is good under normal circumstances and that it should be open so that the identity of the peer reviewers is known.

4.4. Feedback on access to research data and speed of COVID research

Interviewee 1 noted that data sharing was not good during the pandemic and that many countries had limitations on sharing tissues, fluids, and clinical information due to restrictions aimed at protecting their citizens. However, he also shares an example of a group at Harvard that was willing to share data analysis, RNA-SEQ, and antibody panels. Interviewee 1 was impressed with this group's willingness to share their data and believes that they went above and beyond their typical responsibilities to do so. He suggests that while there were both good and bad examples of data sharing during the pandemic, there is room for improvement in terms of sharing clinical data for all groups.

Interviewee 2 stressed the importance of open data and sharing data during the COVID-19 pandemic, particularly with regards to the SARS-CoV-2 genome. The interviewee highlighted the role of GISAID, an open-access platform for sharing genomic data, and emphasised the importance of acknowledging the author who submitted the genomic sequence of SARS-CoV-2 from their lab. The interviewee noted that while he agrees with open data sharing, it is important to acknowledge every author who submitted their data, and that GISAID protects the rights of authors who submit their data to the platform. The interviewee also mentioned the importance of equity in the sharing of data, particularly with developing countries, and outlined how GISAID was established due to the avian influenza virus in Indonesia.

Interviewee 3 believes that the COVID pandemic has contributed towards more data sharing; he is involved in EU-funded projects that tries to make COVID-related data available in the open science cloud as quickly as possible. He acknowledges that there are debates about which data are relevant and what standards need to be met, but he has not had any issues with accessing COVID-related data, as there are portals available, and it is easy to find the data he needs.

Interviewee 4 used mostly government data for her research but found that the quality of the data was not always good – but it was readily accessible. She considers sharing to be still inadequate, even when journals require it: some published articles with data sharing statements have links that do not work or have incomplete data. The interviewee believes that statistical analysis should be supported by data that is made available, and that data should be available by the time of publication.

Interviewee 5 stressed the importance of data sharing in COVID-related research, particularly the need to provide access to all the data that is published to allow for verifiability and deeper interpretation. He mentions that all data related to COVID in their published papers were uploaded to supplement files. He cautions that raw data needs to be given meaning, and interpretation is still necessary. He believes that data sharing can help with the verification process and that open access to data could allow for more efficient research.

Interviewee 6 is part of a network funded by the EU, which aims to exchange real-time data from hospitals and the healthcare system. He thinks that working with original data requires full access, which may be difficult due to language barriers and GDPR issues. The interviewee hopes that access to data will be easier for the next pandemic. He suggests that those who claim not to have had a problem probably only did literature reviews and did not work with original data. The interviewee mentioned that the pandemic has highlighted legal issues related to data protection and sharing across country borders, and there is a need for a common legal framework within Europe to facilitate easier and faster exchange of clinical data.

4.5. Feedback on data quality

Interviewee 1 noted that due to the newness of COVID-19 and the lack of a historical context, it was difficult to distinguish good data from bad data. He believes that while some of the data related to COVID-19 is not useable, some of it is good. In response to the question of whether COVID-19 will result in more data sharing, interviewee 1 is not optimistic, noting that data sharing involves more than just sharing a piece of paper, but also includes samples, clinical information,

and behaviour information and samples, and that there is a significant knowledge gap in some areas, such as MRI data. He believes that while there has been some progress in data sharing, he was expecting more from the response to COVID-19.

Interviewee 2 described the quality control process for submitting SARS-CoV-2 genome sequence data to GISAID. The quality of the data is assessed by checking the end stretch, and if it is more than 30%, the data cannot be submitted. The speaker also mentioned that the resources and the quality of the RNA and the samples themselves need to be taken into account when assessing the quality of the data. If the data is very bad, the speaker suggested checking the sequencing failure rate, and if it is less than 30, the data can be used. The data is also checked by GISAID to ensure it is real data before it is released.

Interviewee 3 did not recall any bad papers but thinks that it's important to set standards for certain sets of data. He notes that there are efforts in his research field to set up standards for in vitro asset data for developing small molecule therapeutics and believe this is an important development that needs to happen.

Interviewee 4 pointed to poor quality data that is sometimes provided by governments and its impact on research studies. She explained that sometimes the data was poorly defined, or the definitions changed midway, making it difficult for researchers to use the data for their studies. She also mentioned an example of a research study that was taken out of context, and its data was misused because it was not clearly labelled. The study had used many times more COVID-19 particles than what would normally be expected, and the results were used to support measures such as plexiglass and bleach, which were not necessary outside of a hospital setting.

Interviewee 5 pointed to the challenges of conducting a meta-analysis on COVID research due to the lack of controlled studies with the same groups. He suggests that the publication of position papers with minimum requirements for data monitoring and comparison would be helpful in fields like immuno-monitoring of vaccine efficacy. The speaker also mentions that neutralizing antibodies are not always effective in COVID, and that additional data should be involved and monitored to approach this problem.

Interviewee 6 encountered bad quality data, but due to the nature of the data, originating from ICUs, mistakes can happen when physicians and nurses record data into the system. Having enough data is crucial to get a statistical power to understand the trends in the data so as to have conclusive data. As his country has a small population, they were dependent on other European data to get statistical power to make scientifically sound conclusions.

4.6. Feedback on importance of access to publications vis-à-vis access to data

The number of survey respondents that considered sharing publications an important factor for the speed of COVID research is higher than the number of people that put the same importance on sharing of research data (see above)

Interviewee 1 believes that this stems from the fact that the title of the publication can be exciting. Furthermore, the data itself is often not of sufficiently available or of sufficiently high quality to be checked so one makes do with the publication. The interviewee also discusses the importance of transparent peer review, with Bio and eLife being leaders in this area. The interviewee believes that the rewards for peer review, such as recognition and points for Starbucks coffee, are unnecessary, as peer review is part of a scientist's job, and the success of one scientist is a success for all scientists.

Interviewee 2 was not asked this question due to time constraints.

Interviewee 3 stated that it might be difficult for people to understand and appreciate data without having metadata that provides a real understanding of how the data were generated, the protocols used, and the purpose of the research. People therefore prefer to look at publications that provide a proper introduction and a description of the results, as well as the data as a whole package, which makes more sense to them. He suggests that not everyone may enjoy looking into databases and pulling out data without context.

Interviewee 4 believes that many COVID-19 research authors are not doing primary analysis and are instead summarizing other people's work. For some research studies, access to data is not essential, such as for a basic review that summarizes previous papers. She explains that in her paper where she reviewed 188 papers, she did not need access to the data but rather just the conclusions. In another article, where the double-checked the statistical analysis of a paper, access to data was essential. She notes that many authors from COVID-19 weren't actually looking at data, and if one is not looking at data, then the availability of data doesn't matter.

Interviewee 5 emphasised the importance of interpretation over raw data and the value of storytelling in understanding the purpose and context of research. Both publications and data are relevant, but it can be helpful to hear the narrative first.

Interviewee 6 explained that during the pandemic, many people who were not originally virologists or epidemiologists started working in those fields, and they were more dependent on published data than raw data. It was logical for them to rely on publications since they were not in their own scientific field.

4.7. Feedback on other open science issues and the speed of COVID research

Interviewee 1 was not specifically asked this question.

Interviewee 2 emphasised that it's important for researchers to acknowledge and give credit to the work of others whose data they use. The interviewee also mentions that in some cases, researchers in developing countries have limited resources and may therefore not be able to perform advanced research, so sharing data can help them do their job and make progress in their research. He suggests that changing the culture around acknowledging data sharing could benefit researchers who have limited resources.

Interviewee 3 thinks that it's not a good idea to solely judge the work of scientists based on the number of publications or where their work is published (i.e., first author, last author, or in a prestigious journal). He notes that people may prefer to publish in open science journals rather than traditional ones, and that the traditional system needs to change. He does not have an answer as to what the best assessment criteria would be but suggest that the number of citations in open science publications could be useful. He believes that the traditional way of assessing scientific work is no longer effective.

Interviewee 4 pointed to US legislation which requires NIH-funded studies to become open access within a shorter period of time, and the data must be open at the same time. The author suggests that a private market solution could be a better option, with professional organization journals developing a new system for paying their basic costs, which would not require authors to pay a fee or be funded by advertising. The author also believes that online advertising could be a way to pay for publication costs and reduce the influence of big pharma and grant money.

Interviewee 5 points to the fact that open access is often required by funders, including in his own country, and criticises the need to pay to publish because high APCs can act as a filter. He suggests that both open access and traditional publishing systems should be available and criticises the unequal distribution of wealth which also affects scientists.

Interviewee 6 believes that most of the data produced or funded by public sources, hospitals, or universities should be freely available and exchanged on a worldwide basis as this is how science works. The interviewee calls for open data exchange and rejects practices like those in North Korea where COVID cases are not reported. He suggests improving data exchange in the scientific community.

4.8. Feedback on the effectiveness of the scientific system in tackling COVID

Interviewee 1 believes that the current system of data sharing is not working and that most labs are not sharing data. He thinks that people are in a "comfy spot" where they claim they are willing to share data but do not actually do so (with his lab being one of the exceptions). He also points out that there are issues related to patient data protection. Interviewee 1 believes that in Europe, targeting the head of research groups for sharing data could be a smart idea as they have the power to establish collaborations as they make decisions for their group.

Interviewee 2 holds that the collaboration and sharing of data among researchers played a significant role in COVID successes, as well as the faster pace of research due to open data. The interviewee also mentions that the research achievements related to COVID-19 are more prominent and successful compared to other viruses like dengue, where a successful vaccine has yet to be developed. He highlights that there is room for improvement, but the successful stories in COVID-19 research should be acknowledged and celebrated.

Interviewee 3 acknowledged that there are many COVID-related publications due to the topic's popularity and funding opportunities. He would like to see more openness with clinical trial data sponsored by pharmaceutical companies. He believes in open science and open data but recognises that there may need to be compromises to encourage scientists to work. He suggests that the issues of publications, data, and research assessments are the most important aspects to focus on when considering the importance of open science in the speed of COVID research.

Interviewee 4 expresses a negative opinion about the scientific community's handling of data during the COVID-19 pandemic, calling it the "biggest data disaster in the history of the world." She thinks that much data has been withheld, such as preliminary data from vaccine trials not being made immediately available, and good articles having been withdrawn without being published. She also criticises the peer review process and academic culture, stating that it is a "crony system" and broken.

Interviewee 5 believes that the scientific system worked well during COVID-19, thanks to the collaboration between researchers, medical doctors, and funding agencies. He emphasises the importance of interest and engagement in the subject matter to achieve success. The interviewee also notes that close collaborations with medical doctors were previously lacking but were improved during the pandemic.

Interviewee 6 stated that the scientific system produced a well-working vaccine in a short time, showing the power of science. However, there is a lot to learn in terms of communication to the public. The interviewee believes that progress has been made in this area, but there is still room for improvement.

He believes that both gatekeepers to validate research and open communication that is easily understood by the public are needed in the scientific system. The interviewee mentioned that extreme opinions in science communication can be confusing for people and that scientists need to follow quality rules in science, including peer review, publications, and open discussions.

4.9. Final thoughts on open science and the speed of COVID research

Interviewee 1 reflected on the positive and negative aspects of science during the COVID-19 pandemic. He believes it exposed both the best and worst of the scientific community. However, he also believes that the pandemic presented opportunities for collaboration and the potential for future scientific advancements. The interviewee suggests that the efforts made for COVID-19 could be applied to other diseases, such as malaria, Ebola and dengue, to make a positive impact on humanity. He emphasises his belief in the responsibility and importance of doing science and sees potential for positive change in the future.

Interviewee 2 was not asked the question due to time constraints.

Interviewee 3 spoke of his experience as being part of a large COVID-19 EU-funded grant that took nearly a year to get started due to security issues related to researchers distributing infectious material. He suggests that, in the event of a similar emergency, it is important to speed up the management processes behind funding and get people working more quickly.

Interviewee 4 suggested quick reviews as being helpful. Open access has for her both advantages and disadvantages: while it provides an advantage to readers, it provides an equally serious disadvantage to authors. She believes that both sides need to be considered. She suggests that pre-print services and platforms like ResearchGate will become more popular, allowing people to publish their work, and allowing the community to judge and determine accuracy. The interviewee believes that the community can judge the reliability of the work based on comments and the reputation of the experts providing them.

Interviewee 5 did not have additional input beyond what was covered in the interview.

Interviewee 6 did not have additional input beyond what was covered in the interview.

4.10. Conclusions from the interviews

The interviewees mentioned several **factors** influencing the speed of COVID research. Interviewee 1 highlighted the influence of political factors in the approval process of the COVID-19 vaccine and the need to balance speed with maintaining scientific standards. Interviewee 2 emphasised the urgency of finding solutions to the lack of available treatments and vaccines for a new disease that has affected almost every country in the world.

Interviewee 3 suggested that there has not been a significant increase in *new* worldwide collaborations for COVID research, but – in his case - rather a focus on accessing new data on potential therapeutics from academic institutes he is already collaborating with. Interviewee 4 stressed the importance of worldwide collaboration, with researchers from various fields coming together to make sense of "chaotic and conflicting data" that has emerged during the pandemic. She also noted that in her view poor quality articles and political considerations influenced what was published during the pandemic.

Interviewee 5's response highlighted the personal desire to help. He mentioned the importance of open sources but also the potential for misleading information, indicating the need for careful monitoring of scientific literature during the pandemic. Interviewee 6 confirmed the importance of international collaboration, with existing networks allowing for sharing experiences, publishing, and establishing research groups.

More specially as regards the **role of open access to scientific publications**, none of the interviewees experienced access problems. Interviewee 1 expresses concern about the potential trade-off between speed and scientific rigor in publications. He also identified limitations to accessing scientific literature in certain contexts, including colleagues in Pakistan and clinicians in developed countries who work in hospitals that are not affiliated with universities. Interviewee 6 also experienced colleagues having access problems.

While most interviewees were positive about open access, interviewee 4 considers Open Access a "mixed bag" since it is not always free for authors. Similarly, interviewee 5 highlighted that open access can be financially challenging for independent researchers.

Overall, the responses suggest that open access publications have been critical for the development of knowledge around COVID-19 during the pandemic. However, there are still limitations in accessing scientific literature in some contexts, which can create inequities in the dissemination of knowledge.

The interviewees had different opinions on the **value of pre-prints, the peer-review process, and the role of speed versus quality in scientific research** during the COVID-19 pandemic.

Interviewee 1 expressed concern about the risk of misinformation and suggests more rigorous review processes to avoid this. This was also partially reflected in the response of interviewee 5, who highlighted the need to monitor the groups publishing pre-prints to be able to accurately ascertain their validity. Interviewees 2, 3 and 4 generally appreciated the role of pre-prints in disseminating information quickly. Interviewee 3 was positive about pre-prints as long as it is clear that the data is to be interpreted "at own risk" and interviewee 4 even preferred pre-prints to peer reviewed publications.

In this context, the interviewees also expressed different views on the peer-review process, with interviewee 1 suggesting a need for more rigorous review processes. Interviewees 5 suggested paying reviewers to speed up the process. Interviewee 6 stressed the need to critically review any kind of information, whether the source has been peer-reviewed or not. In this sense he considers all papers pre-prints.

Overall, while the interviewees have varying opinions on these topics, they all acknowledge the importance of ensuring that scientific research during the pandemic is reliable and properly vetted. This suggests that a balance needs to be struck between the speed of publication and maintaining scientific standards to ensure reliable information is disseminated during pandemics.

As concerns **access to research data** interviewees 1, 4, and 6 expressed concerns about inadequate data sharing during the pandemic, with interviewee 1 stating that while there is room for improvement in sharing clinical data there are also some good practice examples. Interviewee 4 used mostly government data for their research but found that the quality of the data was not always good, and that sharing was still inadequate. Interviewee 6 highlighted legal issues related to data protection and sharing across country borders. In contrast, interviewees 2, 3, and 5 had a more positive outlook on data sharing, with interviewee 2 also emphasizing equity in data sharing. Interviewee 3 believed that the pandemic had pushed towards more data sharing but acknowledged that debates around relevance and standards still exist. Interviewee 5 stressed the importance of data sharing for verifiability and interpretation and noted the potential benefits for future research. Overall, the interviewees agreed that data sharing was crucial for COVID-related research, but there was room for improvement in terms of accessibility, quality, and legal frameworks for sharing across borders.

When specifically asked about **data quality** interviewee 1 noted that due to the newness of COVID-19 and the lack of a historical context, it was difficult to distinguish good data from bad data. He believes that while some of the data related to COVID-19 is “garbage”, some of it is good. In response to the question of whether COVID-19 will result in more data sharing, interviewee 1 is not optimistic, noting that data sharing involves more than just sharing a piece of paper, but also includes samples, clinical information, and behaviour information and samples, and that there is a significant knowledge gap in some areas. While there has been some progress in data sharing, he was expecting more from the response to COVID-19.

Interviewee 2 was more positive about the quality of COVID data, describing the quality control process for submitting SARS-CoV-2 genome sequence data to GISAID. Interviewee 3 did not recall any particularly bad papers but thinks that it's important to set standards for certain sets of data. He notes that there are efforts in his research field to set up standards for in vitro asset data for developing small molecule therapeutics and believe this is an important development that needs to happen.

Interviewee 4 discussed poor-quality data that is sometimes provided by governments and its impact on research studies. She explained that sometimes the data was poorly defined, or the definitions changed midway, making it difficult for researchers to use the data for their studies.

Interviewee 5 pointed to the challenges of conducting a meta-analysis on COVID research due to the lack of controlled studies with the same groups. He suggests that the publication of position papers with minimum requirements for data monitoring and comparison would be helpful.

Interviewee 6 encountered bad quality data, but due to the nature of the data, originating from ICUs, mistakes can happen when physicians and nurses record data into the system. Having enough data is crucial to get a statistical power so as to have conclusive data.

The interviewees also provided varied perspectives on the **relationship between data and publications** and the fact that survey respondents ranked access to publications as more important than access to data for the speed of COVID research.

Interviewee 1 argued that lack of (high quality) data may be to blame, while interviewees 3 and 5 highlighted the importance of context (metadata) and interpretation/storytelling in understanding results, and this is often provided in the publication. Interviewee 4 suggested that access to data is not essential for all research studies, only those that involve statistical analysis. Interviewee 6 explained that during the pandemic, many people outside of virology and epidemiology started working in those fields, making publications more important than datasets.

In regard to **other open science issues** and the speed of COVID research the interviewees discussed the importance of data sharing, acknowledging the work of others, and the need for a shift in the traditional scientific publication model. Interviewee 2 pointed to the need to changing the culture around acknowledging data sharing which could benefit researchers with limited resources, in particular researchers in developing countries.

Interviewee 3 noted that the traditional system of assessing scientific work based on the number of publications or the prestige of the journal needs to change, and that open science journals may be a useful alternative. Interviewee 4 pointed to US legislation requiring NIH-funded studies to become open access, and suggested a private market solution to fund journal costs without relying on author fees or advertising. Interviewee 5 highlights the issue of high article processing charges (APCs) and calls for both open access and traditional publishing systems to be available to reduce the impact of wealth inequality on scientists. Interviewee 6 called for better data exchange practices in the scientific community. Overall, the interviewees highlight the need for more open access and transparent scientific practices, as well as more equitable distribution of resources and opportunities for researchers worldwide.

As regards the **effectiveness of the scientific system in tackling COVID,** reactions were mixed: interviewee 1 believed that the current system of data sharing is not working, with most labs not sharing data, and suggests targeting the head of research groups in Europe to establish collaborations. Interviewee 4 also held a negative opinion of the scientific community's handling of data and censorship during the pandemic, calling it the biggest data disaster in history. In contrast, interviewee 2 highlighted the successful collaboration and sharing of data in COVID-19 research and suggests celebrating these achievements, while acknowledging that there is still room for improvement. Interviewee 3 called for more openness with clinical trial data sponsored by pharmaceutical companies; he believes in open science and open data but recognises that compromises may be needed. Interviewee 5 also thought that the scientific system worked well during COVID thanks to the close collaboration of the main stakeholders. Interviewee 6 praised the scientific system's success in producing a vaccine in a short time but emphasises the need for both gatekeepers to validate research *and* open communication easily understood by the public.

As regards their final thoughts on open science and the speed of COVID research some of the interviewees provided a range of reflections. Interviewee 1 saw both positive and negative aspects of the scientific community during the pandemic and believes that future scientific advancements can be made by applying the efforts made for COVID-19 to other diseases. Interviewee 3 highlights the need to speed up the management processes behind funding in the event of a similar emergency. Interviewee 4 suggests quick reviews as being helpful and suggests that pre-print

services and platforms will become more popular, allowing people to publish their work, and allowing the community to judge and determine accuracy.

5. Overall Study Conclusions and Policy Implications

Open access to relevant publications and data has been depicted as a key asset in combatting the COVID-19 pandemic (e.g. Besançon et al 2021[11], OECD 2020[12], Sparc Europe[13] w.d.) In January 2020 a number of organisations, including publishers, funders, and scientific societies signed a call to ensure that research findings and data relevant to this outbreak are were rapidly and openly to inform the public health response and help save lives, sometimes referred to as the "Wellcome statement"[14] or "Joint Statement". The US Executive Order on Ensuring a Data-Driven Response to COVID-19 and Future High-Consequence Public Health Threats of January 21, 2021 also included a commitment to openness.[15] Furthermore, the European Union launched a flurry of COVID initiatives, including emergency funding calls with stringent openness requirements, the ERA corona platform[16] and a COVID-19 data portal[17]. EU COVID calls also contained stringent openness requirements. Additional actors such as the EOSC secretariat explicitly link open science to the speed of discovery (see literature review).

Our research question whether and to what extent Open Science contributed to the speed of COVID research therefore has important policy implications. A few other studies have also looked at aspects of this question, although with slightly different emphasis. **Frontiers' "The Academic Response to COVID-19"[18]** [81] which is based on a survey of 25,307 members of the academic community is not primarily concerned with Open Science and the speed of research but finds that the pandemic has made researchers more likely to publish open access, share their data and use preprint servers, with a higher propensity for open science practices in the global South than in the global North.

A second important document is the report **"Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic"[19] from the Research on Research Institute (RoRI),** which was delivered in their capacity as

[11] https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/s12874-021-01304-y.pdf
[12] https://www.oecd.org/coronavirus/policy-responses/why-open-science-is-critical-to-combatting-covid-19-cd6ab2f9/
[13] https://sparceurope.org/covid-19-and-open-science/
[14] See https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak
[15] See https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/21/executive-order-ensuring-a-data-driven-response-to-covid-19-and-future-high-consequence-public-health-threats/
[16] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/covid-19
[17] https://www.covid19dataportal.org/
[18] https://www.frontiersin.org/articles/10.3389/fpubh.2020.621563/full see chapter 4.
[19] Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic (figshare.com)

scientific advisors to the C19 rapid review group, which in turn was formed in direct response to the Wellcome statement. The report aims to investigate whether the commitments in the Welcome statement were met. In a similar vein, the study "**From intent to impact: Investigating the effects of open sharing commitments"**[20] commissioned by Wellcome, UK Research and Innovation and the Bill & Melinda Gates Foundation, explores the impact of calls to rapidly and openly share COVID-19 research findings to inform the public health response. Both these documents are clad in rather positive language, especially regarding the success in opening up COVID literature but also highlight some weaknesses and challenges, in as regards pre-prints and data sharing practices. These points are also taken up by the OECD's Going Digital Toolkit Note "Open Science - Enabling Discovery in the Digital Age".[21]

How does our study fit in here? In many ways we confirm some of these earlier finding. We find access to scientific publications was considered a very important contributing factor to the speed of COVID research, with a majority of respondents satisfied with access to COVID publications. However, in the qualitative interviewees some speakers stressed there are still limitations in accessing scientific literature in some contexts, which can create inequities in the dissemination of knowledge.

Additionally, we also find both in the literature review and the interviews a somewhat more ambiguous attitude to the value of pre-prints, the current state of the peer-review process, and the role of speed versus quality in scientific research during the COVID-19 among the scientific community than would have been expected if one would simply read policy makers' statements.

Our findings by and large share the consensus of the earlier studies quoted above as concerns research data sharing, namely that while it did play a role in the speed of COVID, this was less pronounced than with open access to publications. In our survey, only a narrow majority was satisfied with the quality of available COVID data. This was also reflected in the interview phase: interviewees agreed that data sharing was crucial for COVID-related research, but many pointed to weaknesses in terms of accessibility, quality, and legal frameworks for sharing across borders. Several interviewees also highlighted the need to change the culture around acknowledging data sharing and the traditional system of assessing scientific work based on the number of publications or the prestige of the journal.

All in all, it is clear that COVID has acted as a "booster" for open science and several authors and interviewees have expressed the hope that measures to increase the openness of research can also be applied to other infectious diseases, medical conditions, and beyond. At the same time, the key question is whether this "COVID booster" for open science has been temporary or whether new open practices will form a new baseline on the road to a scientific system were Open Science is "the new normal". Generally, I stand by my conclusion from my pre-COVID 2020 article in the LSE impact blog in which I argued that Open Science is not yet the default setting.[22] At the time I

[20] https://zenodo.org/record/7003684#.ZA-PNXbMIxV
[21] https://goingdigital.oecd.org/data/notes/No13_ToolkitNote_OpenScience.pdf
[22] https://blogs.lse.ac.uk/impactofsocialsciences/2020/01/17/not-yet-the-default-setting-in-2020-open-research-remains-a-work-in-progress/ see also a 2023 update on https://www.linkedin.com/pulse/open-science-why-arent-we-yet-daniel-spichtinger/

argued that we face a number of overarching obstacles which we have to address if we want to continue to make progress. One is overcoming the resistance of actors with significant vested interest in the status quo. Another concern is geopolitics and the instinct to hoard data; since 2020 the global context for open science has only got worse. This includes barriers like mandatory geolocation and data sovereignty but also (purposefully?) divergent legal frameworks which make data exchange difficult. Third is the need to change research culture.[23]

The results from this study, in particular the intrinsic linkage (co-dependency) of open access and open data to wider open science issues and their controversial discussion among the scientific community, shows a science system that is still in transition and flux. In other words: while the actions to open knowledge taken during the pandemic (such as the Joint Statement) have given open science a boost, it is too early to declare the open science mission as accomplished. Ensuring that we advance further on the road to Open Science and that the boost to Open Science becomes the new baseline and not a temporary phenomenon also means that we need to be vigilant against backsliding (such as publishers closing articles that were made open during COVID).

While the road to open science may be a case of **"two steps forward, one step back**" it is important that we continue to make progress in the post-COVID era, for example through initiatives such as the coalition to reform research assessment.[24] However, we will not achieve open science through one silver bullet activity; rather we need to patiently continue on the long and winding road to open science. There are a number of ongoing activities we need to continue or expand to further advance the open science cause, such as (i) continuous awareness raising and training activities among researchers, (ii) ensuring we have robust mechanisms for monitoring the state of play in place, (iii) having effective compliance mechanisms and sanctions from funders as well as (iv) incentives and (v) support mechanisms.[25] All in all, individual initiatives in different areas of open science are very necessary and valuable to give us a boost. But in the end implementing open science is a marathon, not a sprint.

6. Annexes

The following separate annexes are available with supplementary data:

- **Annex 1: COVID OS Speed Literature review** (excel file, contains google search and google scholar search)
- **Annex 2: Interview Consent and Information Form**
- **Annex 3: Interview Guide**

[23] See for instance https://www.csescienceeditor.org/article/toward-open-science-contributing-to-research-culture-change/
[24] https://coara.eu/
[25] In this context, it is significant that in a survey of 3144 researchers 31% saw increased demands for open science without extra resources as a challenge (in 4th place of all challenges). See Economist Impact/Elsevier (2022)Confidence in research: researchers in the spotlight. P.34. https://impact.economist.com/projects/confidence-in-research/