Science-Metrix

Understanding the impact of the COVID-19 open sharing statement

Annex A: Bibliometric technical report

Date: June 15, 2022

With contributions from Etienne Vignola-Gagné, David Campbell, Jeroen Baas, Elisabeth Browning, Attila Emecz, Stephanie Faulkner, Daniel Fernandez-Guarda, Paul Khayat, Michael Magoulias, Henrique Pinheiro, Simon Provençal, Celina Sprague, Seyedamin Tabatabaei, Georgios Tsatsaronis, Nicolien van der Linden, Jan-Maarten van Schalkwijk (Science-Metrix/Elsevier)

Euan Adie (Overton)

Andrea Chiarelli, Lucia Loffreda, Rob Johnson (Research Consulting)

Submitted to:
Wellcome, UKRI, and the Bill & Melinda Gates Foundation

By:

Science-Metrix, an Elsevier company
www.science-metrix.com

Contents

Tables

Figures

Acronyms and short versions

Big5	The group of the five largest academic publishing houses: Elsevier, Sage, Springer Nature, Taylor & Francis, and Wiley/Wiley-Blackwell
DAS	Data availability statement
DiD	Difference-in-differences
HVRD	Human viral respiratory diseases
HVVD	Human viral vector-borne diseases
Joint Statement	the Joint Statement on *Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak*
OA	Open Access
RPO	Research performing organisation
Sig.	Signatory, signatory organization to the Joint Statement
ToC	Theory of change

1 Introduction

Following early signs of what would become the global COVID-19 pandemic (in January 2020), Wellcome, the Bill and Melinda Gates Foundation and UK Research and Innovation published a Statement (henceforth the 'Joint Statement') calling on researchers, journals, and funders to 'ensure that research findings and data relevant to this outbreak are shared rapidly and openly to inform the public health response and help save lives'. To support the main sponsors and instigators of the Joint Statement in their strategic and organisational planning of future open science activities, Research Consulting and Science-Metrix co-designed a study that attempts to isolate and quantify the specific effects of Joint Statement endorsement on COVID-19 research benefitting from support by the signatory organisations, or published with signatory journals. In particular, the study was designed as a mixed-methods approach (desk research, interviews, survey, and bibliometrics) aiming to capture and triangulate evidence as it relates to the Joint Statement's outputs and outcomes towards realising open sciences principles, and most notably, for fostering rapid sharing of research findings and underlying data sets. The current report provides a comprehensive description of the bibliometric component implemented by Science-Metrix, complementing the findings and methodological overview provided in the main report.[1]

There has been some research conducted in the last two years to try and measure the degree of realisation of data sharing and open science principles in **COVID-19 research overall.** Most bibliometric, informetric or quantitative research has characterised the COVID-19 literature broadly in relation to preprinting, data sharing, OA status or a combination of these dimensions.[2] The Research on Research Institute report titled *Scholarly Communication in Times of Crisis* has examined OA and preprinting practices for COVID-19 research overall, and in more specific analyses, practices related to publishers that are signatories to the COVID-19 Rapid Review Initiative.[3] Funders and investigative science journalists have also reported on the issue based on qualitative or administrative data.[4] As is made clear by this very brief review of prior, directly relevant results, no other study has attempted to specifically measure the differential value of Joint Statement endorsement in signatory research against non-signatory research.

---

[1] Research Consulting, & Science-Metrix. (2022). *From intent to impact: Investigating the effects of open sharing commitments*. Nottingham and Montréal. Commissioned by Wellcome, UK Research and Innovation and the Bill & Melinda Gates Foundation. Retrieved from https://zenodo.org/communities/data-sharing-in-public-health-emergencies/?page=1&size=20.

[2] Collins, A., & Alexander, R. (2022). Reproducibility of COVID-19 pre-prints *. *ArXiv*. Retrieved from https://github.com/anniecollins/reproducibility_markers_in_covid19_preprints.; Fraser, N. et al. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, *19*(4), p. e3000959. doi:10.1371/JOURNAL.PBIO.3000959; Larregue, J., Vincent-Lamarre, P., Lebaron, F., & Larivière, V. COVID-19: Where is the data? | Impact of Social Sciences. *LSE Impact Blog*. Retrieved from https://blogs.lse.ac.uk/impactofsocialsciences/2020/11/30/covid-19-where-is-the-data/; Helliwell, J. A. et al. (2020). Global academic response to COVID-19: Cross-sectional study. *Learned Publishing*, *33*(4), pp. 385–393. doi:10.1002/LEAP.1317; Li, R. et al. (2021). COVID-19 trials: declarations of data sharing intentions at trial registration and at publication. *Trials*, *22*(1), pp. 1–5. doi:10.1186/S13063-021-05104-Z/TABLES/3.

[3] Waltman, L. et al. (6 December 2021). *Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic*. Research on Research Institute. doi:10.6084/M9.FIGSHARE.17125394.V1.

[4] Watson, C. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nature Medicine*, *28*(1), pp. 2–5. doi:10.1038/S41591-021-01654-6; NIAID Media Team. Reflections on a Year of COVID-19 Data Sharing | NIH: National Institute of Allergy and Infectious Diseases. *NIAID Now Blog*. Retrieved from https://www.niaid.nih.gov/news-events/year-covid-19-data-sharing.

1.1 Bibliometric approaches to program evaluation and outcomes measurement

To fulfill this mandate, Science-Metrix has decided to treat the release of the Joint Statement as an intervention ultimately aiming to change open science behaviors within a scientific community or system, albeit in this case a particularly large community. The Joint Statement can be treated similarly to a mission-oriented grant that aims to support a group of researchers while at the same time influencing some of their practices to align with policymakers' own specific goals.

Accordingly, a program evaluation approach has been deployed to compare openness and data sharing levels in relevant streams of research (COVID-19 research or the closest comparable research streams prior to the pandemic), before and after the start of the pandemic and the subsequent release of the Joint Statement. To help control for secular trends systematically affecting open sciences practices or even health science research more broadly, the before and after comparison of **signatory** COVID-19 research is itself compared against the before and after comparison of **non-signatory** COVID-19 research.

1.2 Difference-in-differences design used in this study

Comparing a treatment group's performance before, during, and after an intervention (a new programming orientation; a new grant; a new collaboration mechanism; a new instrument or support mechanism to support goals such as gender equity or cross-disciplinarity) provides a measurement of the maximum potential difference brought about by the intervention. Benchmarking against comparable groups that did not receive the intervention of interest but that would have been affected by the same or similar local and global trends allows analysts to isolate the precise magnitude of changes introduced by an intervention of interest. This is often referred to as a difference-in-differences (DiD) approach and is commonly used in program evaluations.[5]

Control groups should be selected from suitable external comparators, with comparators ideally sharing several features with signatory COVID-19 research. Some of the factors that can be expected to differentially affect open science practices include

- Disciplinary subfield of publications, with the view that open science practices may vary in prevalence and character by subfield, much like other publication practices;
- Seniority of researchers;
- Gender of researchers;
- Authority, prestige, and/or resource availability of organisations enabling or supporting the research;
- Country of origin of the research, determining socio-cultural contexts of the research but also access to financial and infrastructural support for research.

[5] Langfeldt, L., & Scordato, L. (2015). *Assessing the broader impacts of research. A review of methods and practices*. Oslo: Nordic Institute for Studies in Innovation, Research and Education. Retrieved from https://nifu.brage.unit.no/nifu-xmlui/handle/11250/282742.

By comparing four groups[6] that are as similar as possible on such dimensions, DiD strategies are able to control for certain external factors. Otherwise, signal on changes in performance could possibly be tracked back to the influence of one of the characteristics in the intervention group, rather than the effects of the intervention itself.

1.3 Signatory and non-signatory groups

Capturing the specific impacts and outcomes of the Joint Statement entailed distinguishing between COVID-19 research outputs (i.e. journal publications as indexed in Scopus and preprints as indexed in medRxiv, bioRxiv, arXiv, and SSRN) associated with signatory (intervention) and non-signatory (control) organisations while maximising their comparability per the above DiD requirements. But even within the signatory category, different layers of engagement with the Joint Statement were afforded by the fact that signatory organisations included different sectors of the science system, each contributing differently to the knowledge process. Signatory organisations included higher education institutions, research institutions, funding bodies, professional associations, individual journals, and journal publishers.

Three signatory layers were of particular importance in supporting open sciences practices with COVID-19 research: 1) journals and publishers, 2) funders, and 3) research performing organizations (RPO), which, through affiliation linkages, also relate to individual researchers. A single signatory COVID-19 paper could receive its signatory attribution from all three layers, but also from two or just one. Put differently, the pool of 'signatory COVID-19 publications' could be variously defined as those papers with signatory status on at least one layer, or only papers fulfilling a combination of these three layers.

Below, descriptive findings are provided for all three signatory layers. Furthermore, the journal signatory layer has been further disaggregated into signatory publications from a signatory journal operated by one of the five major scientific publishers; or from a signatory journal owned by a smaller scientific publisher.

Within DiD analyses,[7] it is necessary to precisely define what constitutes groups of signatory publications. Non-signatory publications (control groups) have traditionally been defined as publications without signatory attribution to any of the three layers presented here. Delineating signatory research is more complex, as mentioned above. In the core DiD models used in the study, signatory research was either defined as publications with 'funding- OR journal-based signatory support' or with 'funding- AND journal-based signatory support'. The value of using both definitions is that the first formulation would be expected to have more external validity, whereas the second would be expected to have stronger differential signals but with somewhat less representativity.

The main DiD models fully excluded signatory publications with affiliation-based signatory status (referred to as RPO signatories in the tables). This was done because there were a comparatively low number of publications attributed to this signatory layer (about 3,000 out of 50,000 total signatory journal

[6] A) intervention group, B) prior performances by the intervention group, C) control group D) prior performances by the control group. DiD = (A-B)-(C-D).

[7] The implementation of DiD strategies in the subset of preprint publications did not lead to robust findings, and therefore these findings are not presented here.

publications); and also because, unlike journal- and funding-based signatory status, affiliation-based status was not expected to vary for a single author. This lack of variation at the author level prevented the use of an approach keeping constant the set of contributing authors across the four groups underlying DiD models. This is to maximise comparability across the four groups and thereby control for confounding factors in assessing the likely effects of the Joint Statement.

Attribution of signatory status to publications was based on regex-based queries within funding acknowledgements; the use of curated affiliation metadata from Scopus (for RPO signatory status); or based on journal and publisher information from Scopus. For preprints, attribution of signatory status only relied on regex-based queries within funding acknowledgements, or, in a negligible number of preprints, RPO-based signatory status. Note that because of the format of the metadata and full text obtained from arXiv and SSRN, additional regex queries had to be used to isolate the funding acknowledgement sections of preprints from these sources (see the appendices to this report for further details).

1.4 Prior groups

With most of COVID-19 research at the time of the study having been conducted in 2020 and 2021, a symmetrical period of two years (2018–2019) was employed to define prior groups of research, both for signatory and non-signatory research.

The main challenge in defining prior research groups for use in the DiD comparisons was that COVID-19 research did not exist in 2018–2019 (with the exception of a few publications issued in December 2019). As an alternative to defining groups of thematically comparable publications prior to the COVID-19 pandemic, we designed a thematic data set of publications on 'human viral respiratory diseases' (HVRD) that notably includes research on prior coronaviruses, SARS and MERS, among others. The obvious limitation to this strategy is that HVRD research would be expected to be mostly biomedical research, whereas COVID-19 spurred research of an unprecedented disciplinary diversity (e.g. health, AI, social sciences) as the research community responded to the pandemic. Steps have been taken to control for this potential asymmetry (section 1.7).

1.5 Analytical groups in the assessment of the longevity of Zika Statement-associated outputs and outcomes

Time series were produced for Zika research and a control of comparable research in human vector-borne viral disease (HVVD). Again it was possible to distinguish here between signatory research and non-signatory research, both for Zika and HVVD, resulting in four distinct analytical groups. Note that Zika Statement signatory organizations differ from COVID-19 signatory organizations. The analytical period used was 2014–2020, with an initial intention that the years 2014–2015 provide a baseline measurement against which to assess Zika Statement effects from 2016 onwards. However, the number of Zika journal publications and preprints with workable metadata have turned out to be too low in 2014–2015 to enable this approach.

1.6 Bootstrapping of descriptive and differential findings

The application of standard statistical tests of significance and effect sizes to the large-scale descriptive findings produced by bibliometric exercises is not straightforward and has been the subject of debates in the expert community.[8] Science-Metrix advocates the use of stability intervals to assess uncertainties. Stability intervals inform the uncertainty of bibliometric indicators by providing a range within which a computed score could likely fluctuate in response to a change in the underlying set of publications that was used to compute it. Stability intervals are built by randomly resampling, with replacement, a group's papers to produce many resamples (e.g. N = 2000) of equal size to the group's number of papers. The various indicators to be produced are then computed for each resample to produce an empirical distribution of the scores. This enables the computation of a 95% stability interval—that is, the interval containing 95% of the resamples' scores.

To take a hypothetical example, a stability interval could be used to build the range of scores including 95% of the likely values for the proportion of hybrid OA publications published by two groupings. If the share of hybrid OA papers in group A equals 13% with a 95% stability interval ranging from 11% to 15%, and the share of group B equals 17% with an interval ranging from 16% to 18%, then it would be safe to conclude that group B would perform better than group A even if the underlying sets of publications were to change. The rule of thumb is that if the 95% intervals of the groups being compared do not overlap, then the observed difference is highly likely to remain visible should the underlying data be altered. Because they are built empirically, stability intervals do not rely on the assumptions that the study samples are random and follow a specific distribution. However, they assume that the observed data are representative of the larger populations to which they belong.

1.7 Additional controls and DiD specifications

In the end, six DiD models are presented below, with three different control specifications: 1) all publications, 2) controlled authors, and 3) controlled weighted authors. These three control specifications were combined with two main definitions of signatory status: 1) signatory publications as publications with either signatory funding OR signatory journal status, and 2) signatory publications defined as publications with signatory funding AND signatory journal status.

To simultaneously control for factors such as gender, seniority, and disciplinary background, which may be distributed differently amongst authors and their papers included in the various analytical groups, DiD models have been elaborated where the same authors have been kept constant in all four groups. That is, these models kept only publications written by at least one author having contributed to all four analytical groups of a given model—1) authors that have simultaneously contributed publications to one or more signatory COVID-19 publication(s), 2) to one or more non-signatory COVID-19 publication(s), 3) to one or more signatory HVRD publication(s), and fourthly also to one or more non-signatory HVRD publications An assumption underpinning this specification was that a single author could contribute to

[8] Schneider, J. W. (2015). Null hypothesis significance tests. A mix-up of two different theories: the basis for widespread confusion and numerous misinterpretations. *Scientometrics*, *102*(1), pp. 411–432. doi:10.1007/s11192-014-1251-5.

both signatory and non-signatory research. This specification is presented as 'DiD with controlled authors' below.

In a further development of this approach, a weighting system was developed to ascertain that the variability in publications introduced by the bootstrapping is accomplished while keeping the exact same distribution of authors in the prior and control groups as in the intervention group. This specification is presented as 'DiD with controlled, weighted authors' below.

The rationale for applying the above three control specifications is that while the ability to control for confounders increases from the first to the third one, the size of the resulting convenience samples is increasingly small owing to the more restrictive selection criteria, thereby decreasing the power to uncover statistically significant DiDs. The consistency of results across all three approaches (DiDs pointing in the same direction) was thus used as a sign of robustness in the study's findings even if they were not statistically significant in all cases. Note, however, that the opposite is not indicative of a lack of a positive or negative effect but simply that results are inconsistent and inconclusive.

1.8 Quick overview of selected indicators

The specific changes advocated for in the Joint Statement have been operationalised in five core bibliometrics indicators:

- **Shares (as a percentage) of journal publications (journal articles, reviews and conference papers) publicly available under an open access (OA) licence or free-to-read status** (either directly from the publisher or on an institutional repository). OA and free-to-read status was determined based on the Scopus implementation of the Unpaywall database.
- **Shares of research papers (journal publications and preprints) that include a formal data availability statement (DAS) section**. DAS status has been determined by using regex queries (Appendix C) with a list of commonly found formulations of DAS section headings. DAS status could only be determined for preprints where full-text content was available from source XML files; and for journal publications with full-text archived in Scopus source databases.
- **Shares of research papers that include mention of a data deposition** within their DAS section, with deposits identified when made on an online platform listed in a pre-selected compilation of repositories. Data deposition status of publications has been determined by using regex queries (Appendix D) with a list of the names of open data repositories.
- **Shares of journal publications preceded by a preprint**. The indicator of share of journal publications preceded by a preprint was computed by using a fuzzy matching algorithm to identify authorship, publication year, and title similarities between metadata for the list of COVID-19 and comparable preprints collected, on the one hand; and metadata from the list of relevant journal publications, on the other hand. The matching algorithm was manually validated and improved where necessary.
- **Normalised shares of journal publications with one or more mentions in policy-related documentation, or raw shares of preprints with one or more mentions in policy-related documentation.** Policy-related uptake based on the Overton database, with extra credit to Euan Adie and team for customised text mining of mentions to arXiv preprints; indicator computation

represents share of publications or preprints with one or more citations in a policy-related document (see also Pinheiro et al[9]), following Thelwall's recommendation for caution in dealing with emerging altmetrics indicators.[10] This indicator has been subfield- and year- normalised for findings on journal publications, but it was not possible to apply this strategy to preprints, which are not yet assigned a subfield-level categorisation within the Science-Metrix classification. Normalisation helps control for differences in citation windows for documents published in different years, which may have large impacts on measurements given the fact that policy-related citations normally peak between 3 to 4 years after publication year (see Pinheiro et al). However, comparisons within the preprint set are restricted only to preprints from the year 2020 and 2021, to be mitigated by the analytical period. Caution in the interpretation of policy-related uptake of preprints is nevertheless warranted.

In keeping with the Joint Statement heory of change (ToC) presented in the main report,, signatory COVID-19 research is expected to have experienced differential gains on these five core dimensions against non-signatory COVID-19 research.

1.9 Processing of sources

The relative simplicity of indicator formulations and computations presented above belies the extensive data processing, matching and curation that was necessary to obtain workable metadata and full texts from which to compute the indicators, at scale.

As an illustrative example, Cord19, initially planned as the main source of records of journal publications' and preprints' textual content as part of this study, was discarded early on in the study after it was found that many manuscripts submitted to the database had been stripped of their section headings. This gap would have greatly complicated the identification of DAS or funding acknowledgement sections to clarify signatory status based on financial support.

After a round of initial attempts, the list of sources used was finalised as such:

- Scopus standard metadata records for journal and publisher information of journal publications; funding acknowledgement content for journal publications; author affiliation information for journal publications and preprints; Scopus implementation of Unpaywall for OA and free-to-read status of journal publications.
- Scopus source databases for DAS of publications.
- Selected arXiv records with thematic relevance to the study (identified by initial processing of arXiv's Kaggle metadata database), obtained by targeted, programmatic downloading of arXiv fulltexts. arXiv funding acknowledgements needed to be delineated with a set of regex queries.

[9] Pinheiro, H., Vignola-Gagné, E., & Campbell, D. (2021). A large-scale validation of the relationship between cross-disciplinary research and its uptake in policy-related documents, using the novel Overton altmetrics database. *Quantitative Science Studies*, *2*(2), pp. 616–642. doi:10.1162/qss_a_00137.

[10] Thelwall, M. (2016). *Web indicators for research evaluation: A practical guide*. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 8. Morgan & Claypool Publishers LLC. doi:10.2200/s00733ed1v01y201609icr052.

- bioRxiv and medRxiv metadata and fulltext records obtained from the servers' Amazon S3 buckets in October and November 2021. bioRxiv and medRxiv contained pre-parsed funding acknowledgements sections.
- SSRN metadata and fulltext files obtained from internal Elsevier databases. SSRN funding acknowledgements needed to be delineated with a set of regex queries.
- The PlumX/Science-Metrix implementation of Overton for journal publications and bioRxiv, medRxiv and SSRN preprints. Policy-related mentions toward arXiv preprints were queried using a custom algorithm developed for this project by Overton.

The delineation of this thematic publication set used regex queries for text processing, focusing on isolating COVID-19- or HVRD- associated terms within publications title, abstract and keywords (see Appendix B).

Preprint metadata was examined with a fuzzy logic algorithm to identify journal publications sharing similar first authors, year (or later) of publication, and title, with a view to identify prior preprint versions of publications (an article by Cabanac et al captures a very similar approach to the matching procedure employed[11]). The *Scholarly Communication in Times of Crisis* report has shown that current repositories of preprint-journal publications matches, including those recording databases maintained by preprint servers themselves and recording authors' self-reported final publication versions, were sparse and most certainly lacked many matches. The fuzzy matching approach allowed an expansion of matching coverage while also recording good precision figures in validation against the self-reported matche (94%—see Appendix C).

1.10 Complexity of pandemic-era research and associated limitations

Under normal circumstances, research projects and researchers' publication practices are considered to be relatively stable over multiple years, so that signals captured by a DiD approach can reasonably be traced back to funding programs or other policy interventions. The emergence of pandemic-related research, by contrast, is very likely to be associated with re-shuffling and re-specification of research practices, collaboration networks, and institutional boundaries and definitions that is possibly unprecedented. Against this background, there is a real risk that the differential signal captured through the DiD analyses also captured other factors and trends in health and biomedical research, or in COVID-19 research specifically, that only emerged in 2020–2021.

For instance, in designing the DiDs used to assess the impacts of the Joint Statement, we could not fully control for signatories that have implemented changes in alignment with the statement's objectives prior to their signature (i.e. they would have responded the same way to the pandemic had the statement not been in place). Results from the survey implemented by Research Consulting indicated that a notable number of signatories would conform to this pattern.

While we still think the implemented DiDs help control some of the possible confounders, the approach was most likely not designed to deal with such intense change in such a short period of time. On this

[11] Cabanac, G., Oikonomidi, T., & Boutron, I. (2021). Day-to-day discovery of preprint–publication links. *Scientometrics*, *126*(6), pp. 5285–5304. doi:10.1007/S11192-021-03900-7/FIGURES/6.

basis, we strongly urge the reader to use caution in interpreting the findings that follow and always consider the effects of potential confounding factors.

1.11 Data availability statement

The findings presented in this study can be reproduced by using datasets and code being made available on the project's Zenodo portal at https://zenodo.org/communities/data-sharing-in-public-health-emergencies.

A csv file containing all journal publication and preprint metadata used in the definition of analytical groups and indicator computations is available at 10.5281/zenodo.6582759. Note that snippets of textual context for 1) attributing funding-base signatory status, and 2) for identifying DAS and data deposition mentions have been made available for OA journal publications and for preprints, but could not be made available for other journal publications.

Code to reproduce the indicator computations for empirical measurements (but not the bootstrapping-derived point estimates and stability intervals) from the metadata csv file is also available at the same location.

The full data management plan for the study is available at https://doi.org/10.5281/zenodo.5848642.

2 Results

2.1 Combining and interpreting descriptive, differential and author-controlled findings

For all the indicators provided below, descriptive findings will first be provided to determine whether signatory COVID-19 research has fared better than non-signatory COVID-19 research on various aspects of data sharing and openness in research. Nevertheless, it is not possible from these findings to determine whether higher performance for signatory publications, if any, is at least partly a direct outcome of the Joint Statement's release and its potential implementation within signatory organisations; or whether such high scores would be attributable, in full or partly, to the performances of signatory organisations themselves, which would have been achieved independently of the Joint Statement. Alternatively, increases in the features promoted by the Statement could have prevailed among both the signatory and non-signatory publications, with researchers broadly adopting new research practices for COVID-19 research. Additionally, the qualitative research conducted as part of this project recorded varying levels of Statement-induced organisational change within signatory organisations, providing a further source of uncertainty as to whether signatories' high performances are due to Statement outcomes or to internal policies and practices.

To try and tease out the differential gain specifically introduced by the Joint Statement within signatory organisations' associated performances, the DiD approach compares their observed changes against the backdrop of changes for non-signatory organisations. The pre–post comparison enables the detection of changes that could have been introduced by the Statement while the comparison to the non-signatory organisations controls for the effects of secular trends in, for example, open science policies and practices that might have been triggered by the pandemic across the entire research ecosystem. Therefore, the initial presentation of descriptive findings made below should be interpreted together with the findings of the DiD analyses. Only the DiD results are able to indicate whether good descriptive findings for signatory publications amount to a differential gain against the pre-pandemic baseline, and against a corresponding change within a control group of non-signatory publications.

Recall that this study is still not able to control for new COVID-19 open science policies and practices that have been introduced in signatory organisations in response to the pandemic and might have been adopted fully-independently of the Joint Statement.

2.2 Overview of the COVID-19 publishing landscape and study sample

COVID-19 journal publication and preprint sets, as well as similar sets for control groups of comparable research strands, were collected covering a period extending through October 2021. For the period spanning January 2020 to October 2021, and using the thematic queries presented in Appendix B, we identified slightly more than 160,000 journal publications and slightly more than 38,000 arXiv, bioRxiv, medRxiv and SSRN preprints making up the COVID-19 publication set accessible to this study.

Large shares of these publication sets had to be discarded from the analyses, however, given lack of metadata that would have allowed robust assignment of individual records to the signatory and non-signatory groups. As shown in Figure 1, lack of funding acknowledgement in Scopus records on journal publications was most frequent exclusion condition. Although lack of funding acknowledgement in Scopus can indicate research that has not made use of funding, it was also found that there are instances of clear funding acknowledgement missed by Scopus indexing. In a manual review exercise, slightly more than 20% of journal publications without a funding acknowledgement record were found to have been supported by either signatory or non-signatory funding. More than 40% of records simply had no indication about funding whatsoever. These situations precluded assignment of these journal publications to either the signatory or non-signatory groups. Quite a large number of journal publications were also excluded because they had not been assigned a DOI, or DOI metadata was not found in Scopus (amounting to 4% of the initial publication set).

The study sample was made up of 60,000 journal publications with either signatory or non-signatory status (Table I). Note, however, that non-signatory publications made up 16% of the final sample, while signatory publications made up 84%.

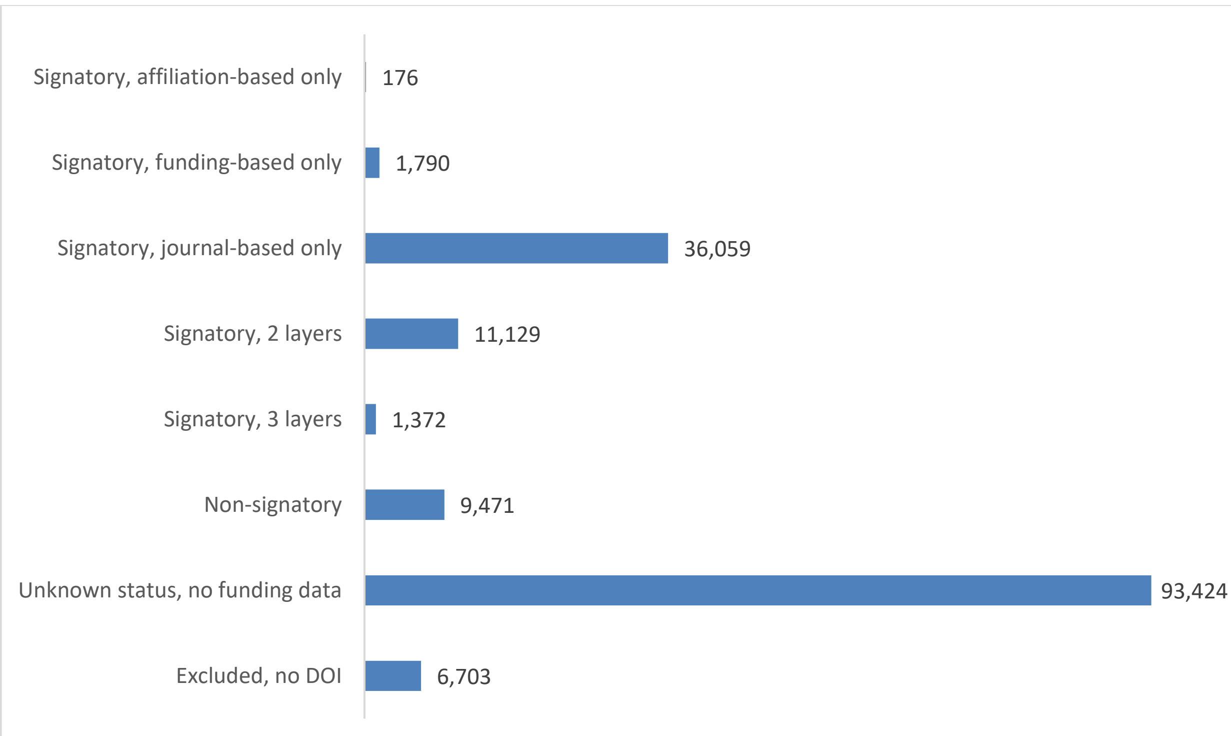

Figure 1 COVID-19 journal publication counts and study analytical group sizes, January 2020 – October 2021

Source: Scopus, processed by Science-Metrix

Table I also provides shares of the overall publications that could be used in the computation of specific indicators. For instance, not all journal publications are covered by the Unpaywall database, and publications that were not covered were fully removed from the analysis. While both Unpaywall and Overton provided good and relatively even coverage of different signatory breakdowns, availability of journal publication records was much lower for DAS writing and data deposition indicators. This is because records of textual content of publications are much scarcer than publication metadata. Note also unavailability of full textual content was skewed towards non-signatory publications, possibly creating some bias in the comparisons between signatory and non-signatory breakdowns on the DAS writing and data deposition indicators. Science-Metrix has not been able to verify the existence of such a bias, quantify it or determine whether it favors signatory publications or non-signatory publications.

An estimate of record availability could not be reliably computed for the indicator of share of journal publications preceded by a preprint. Manual assessment of preprint metadata did reveal variations in the quality of relevant metadata for the matching operation, especially in preprints. Therefore the matching algorithm comparing preprint authors, years and titles to journal publication authors, years and titles was run on all entries, but its success or not in establishing matching would have been limited by differentials in the quality of the metadata for some sets of preprints.

Table I Volumes of COVID-19 journal publications by signatory breakdowns, and by indicator (2020-2021)

Indicator	Non-signatory	RPO signatory	Funding signatory	Overall journal signatory	Non-Big 5 journal signatory
Overall journal publication volume	9,471	2,999	13,038	48,362	14,655
Volume for OA and free-to-read indicators	91%	99%	99%	100%	99%
Volume for preprinting indicator	N/C	N/C	N/C	N/C	N/C
Volume for DAS and data deposition indicators	29%	55%	48%	48%	58%
Volume for uptake in policy-related document indicator	96%	99%	99%	99%	98%

Note: N/C: not computed. See comments in the text. “Non-Big5” : publications published on journals owned by publishing houses outside the Big%. This category is a subset of the “Overall journal signatory” breakdown. Shares of volume by indicator display the proportion of the overall publication records that were specifically used in the computation of that indicator, due to availability of required metadata.

Source: arXiv, bioRxiv, medRxiv, Overton, PlumX, Scopus, SSRN, Scopus, Unpaywall; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

The final set of COVID-19 arXiv, bioRxiv, medRxiv and SSRN preprints had to exclude as much as 87% of the original publication set, due for the most part to the lack of any information about funding or its absence, but sometimes also due to lack of other metadata (see Table II). Following the approach already detailed for journal publications, it was not possible to reliably assign preprints to either signatory or non-signatory groups in the absence of information about funding status. Note that analytical groups are much simpler in the preprint set, since assignment to the signatory group is based solely on funding-based signatory status. Journal-based signatory status does not apply to preprints, and all preprint servers included in this study were signatories to the Joint Statement. Additionally, affiliation-based signatory status had a negligible effect on this component. While 530 preprints were assigned signatory status based on affiliation in the sample, only seven of those did not concomitantly have funding-based signatory assignments.

Note that for preprints (Table II), it was not possible to compute a robust estimate of the proportion of entries where textual content records were complete enough to reliably compute the indicators of DAS writing and data deposition within. While in practice all preprints' records were used for this operation, manual assessment of records has shown that some textual content records were incomplete or were structures in formats that were ill-fitted to regex-based extraction.

Table II Volumes of COVID-19 preprints by signatory breakdowns, and by indicator (2020-2021)

Indicator	Unknown status	Overall non-signatory	Overall signatory	Controlled non-signatory	Controlled signatory
Overall preprint volume	33,102	647	4,373	51	589
Volume for DAS and data deposition indicators	N/C	N/C	N/C	N/C	N/C
Volume for uptake in policy-related document indicator	61%	91%	85%	98%	97%

Note: N/C: not computed. See comments in the text. Shares of volume by indicator display the proportion of the overall publication records that were specifically used in the computation of that indicator, due to availability of required metadata.
Source: arXiv, bioRxiv, medRxiv, Overton, PlumX, Scopus, SSRN, Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

2.3 Open- and free-to-read-access in journal publications

The Joint Statement ToC holds that signatory publishers make journal COVID-19 publications immediately open access or free-to-read. COVID-19 journal publications were therefore examined for their OA or free-to-read status using the Unpaywall data set.

Descriptive findings (Table III) show that between 92% and 96% of COVID-19 journal publications have been made available under an OA modality, depending on the signatory status (through institutional affiliation, funding source, journal of publication; or multiple of these). Within non-signatory publications, the descriptive figure was lower at 82%.

Non-signatory publications recorded much higher shares of gold OA publications (58%) than the three main categories of signatory publications (37% or 38%). However, the finding for signatory articles based on journal attribution (by far the largest group of signatory articles, associated with more than 48,000 records) can be made more nuanced by removing from this breakdown journals owned by the five major scientific publishers, as the distribution of associated journal articles amongst different OA and free-to-read modalities may be different between larger and smaller publishing houses. By restricting the breakdown of signatory publications linked to signatory journals outside the major five publishers, the share of gold OA publications rises to 66% instead, above the corresponding figure for non-signatories. Publications linked to signatory journals outside the major five publishers account for roughly a third (slightly more than 14,500 publications) of all journal-based signatory publications.

The hybrid OA and publisher-based free-to-read categories saw much higher shares of publications in the three main signatory groups than the non-signatory group. For instance, publisher-based free-to-read status was found for 35% of articles with signatory journals, compared to 7% for non-signatory publications. Both figures decreased greatly in the sub-group of journal-based signatory articles from outside the major five publishers (8% for hybrid OA and 7% for publisher-based free-to-read).

Interestingly, the category of repository-based free-to-read publications continues to account for a portion of the COVID-19 literature, at around 5% of non-signatory publications and between 6% and 8% of signatory publications. In the group of signatory publications published by journals not owned by the major five publishers, this share goes up to 16%.

Table III Shares of COVID-19 journal publications available under Open Access or free-to-read modalities, by signatory status (2020–2021)

Indicator	Non-signatory	RPO signatory	Funding signatory	Overall journal signatory	Non-Big 5 journal signatory
Share of overall OA+free-to-read	82.4% [81.6,83.2]	96.3% [95.6,96.9]	93.8% [93.4,94.3]	91.8% [91.5,92.1]	94.6% [94.2,94.9]
Share of gold OA	57.8% [56.7,58.8]	37.5% [35.7,39.2]	37.1% [36.3,37.9]	37.5% [37.1,37.9]	65.5% [64.7,66.3]
Share of hybrid OA	5.6% [5.1,6.1]	19.2% [17.8,20.6]	17.6% [17.0,18.2]	11.9% [11.7,12.2]	7.0% [6.5,7.4]
Share of publisher-based free-to-read (bronze)	14.3% [13.6,15.1]	33.2% [31.5,35.0]	31.4% [30.6,32.2]	35.2% [34.8,35.6]	6.6% [6.2,7.0]
Share of repository-based free-to-read (green)	4.9% [4.4,5.3]	6.4% [5.6,7.3]	7.8% [7.3,8.2]	7.1% [6.9,7.4]	15.5% [14.9,16.1]

Note: Publications for which Unpaywall records were incomplete or missing were considered null and removed from the computation of OA/free-to-read figures. Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: Scopus, Unpaywall; processed by Science-Metrix

Across the various DiD models (from least (left) to most (right) controlled and for the two sets of rules defining signatory status in Table IV), differential scores of OA+free-to-read publishing generally pointed to increases moving from 2018–2019 respiratory viral towards 2020–2021 COVID-19 research. However, the differences in the differential observed for signatory versus non-signatory publications were not consistent across models; in other words, the gains were sometimes smaller and sometimes larger for signatory relative to non-signatory publications and the robust confidence intervals of observed DiDs were often extending on both the positive and negative sides (overlapping with 0). In other words, good signatory scores for OA+free-to-read could not be traced back specifically to the statement.

The reason for this uncertain result appears to be that prior signatory research already recorded high shares of OA and free-to-read publications even before the pandemic, with negative or inconclusive DiDs simply capturing the fact that non-signatory research had possibly more space available to increase its performance on this dimension.

Table IV Differential gains in shares of COVID-19 journal publications available under Open Access or free-to-read modalities, by signatory status (2020–2021)

Indicator	Signatory definition	DiD with all articles	DiD with controlled authors	DiD with controlled, weighted authors
Share of overall OA+free-to-read	fund OR journal	+9.5 p.p. [5.4,13.5]	+4.0 p.p. [-3.2,10.7]	+1.7 p.p. [-7.7,11.2]
	fund AND journal	+1.5 p.p. [-2.6,5.6]	-11.8 p.p. [-23.5,-1.5]	-6.3 p.p. [-19.5,6.4]
Share of gold OA	fund OR journal	-4.5 p.p. [-9.4,0.5]	-3.9 p.p. [-16.1,8.1]	-0.4 p.p. [-16.6,15.6]
	fund AND journal	-10.1 p.p. [-15.2,-4.7]	-14.8 p.p. [-34.6,4.7]	-6.5 p.p. [-30.5,16.3]
Share of hybrid OA	fund OR journal	-2.9 p.p. [-4.9,-0.6]	-3.0 p.p. [-9.4,3.1]	-2.1 p.p. [-10.2,6.4]
	fund AND journal	+0.8 p.p. [-1.9,3.6]	+2.3 p.p. [-9.1,13.2]	-5.0 p.p. [-20.5,11.0]
Share of publisher-based free-to-read (bronze)	fund OR journal	+18.2 p.p. [14.6,21.8]	+11.3 p.p. [2.7,20.7]	+11.4 p.p. [-1.5,25.3]
	fund AND journal	+15.9 p.p. [12.1,19.8]	+7.2 p.p. [-6.4,21.3]	+9.7 p.p. [-11.3,30.9]
Share of repository-based free-to-read (green)	fund OR journal	-1.2 p.p. [-3.8,1.6]	-0.5 p.p. [-6.1,5.6]	-7.0 p.p. [-14.6,1.0]
	fund AND journal	-5.1 p.p. [-8.2,-1.9]	-6.8 p.p. [-15.9,2.2]	-4.0 p.p. [-18.7,8.6]

Note: Publications for which Unpaywall records were incomplete or missing were considered null and removed from the computation of OA/free-to-read figures for this indicator. Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: Scopus, Unpaywall; processed by Science-Metrix

However, the above conclusion was strongly complicated in considering specific categories of OA or free-to-read status. Publisher-managed-free-to-read access (also called 'bronze OA') has greatly increased, from 17% of signatory publications in 2018–2019 to 35% in 2020–2021 leading to a positive DiD, although not always statistically robust, in all models considered.

Share of publications accessible in Gold OA, by contrast, has remained stable between periods for signatory publications. Given that non-signatory publications have seen an increase in Gold OA over the same period, the net result is a differential loss in Gold OA in signatory publications. This result is statistically robust in some but not all formulations of the model and must therefore be interpreted with caution.

In summary, these results point towards overall increases in prevalence of OA and free-to-read publishing during the pandemic generally (without being able to trace this effect specifically to the Joint Statement). Paradoxically, the Joint Statement might have supported publisher-managed-free-to-read access to the detriment of Gold OA, although this conclusion is not yet supported by definitive statistical validation. It can be considered, nevertheless, that out of 48,368 COVID-19 journal articles published with signatory journals, 16,045 were published in journals not owned by the five largest scientific publishers. The general prevalence of publishing with journals owned by the five major scientific publishers may explain the differential increases in publisher-based free-to-read access to publications recorded, which were statistically robust in four out of six models.

2.4 Preprint posting practices

The Joint Statement called for systematic early circulation of COVID-19 findings via preprint servers prior to the submission of manuscripts to journals. This section examines whether signatory research has complied with the preprinting requirement, and, if it hasn't fully, to which extent it has performed better or not than non-signatory research on this aspect. Note that findings on preprinting are subjected to a specific limitation, in that evidence of prior preprinting was collected only from a selection of preprint servers (arXiv, bioRxiv, medRxiv and SSRN). Other preprint servers, such as Research Square, have been shown to be important venues for preprinting of COVID-19 research.[12]

Findings show that preprint posting for journal publications is higher in research on COVID-19 compared to HVRD research. For instance, in the most conservative model, prior preprinting went up from 2% to 11% for signatory articles, and from 1% to 5% for non-signatory articles. In the most controlled model (but covering a much smaller overall set of publications), prior preprinting went up from 5% to 45% for signatory groups, and from 0% to 12% for non-signatory groups.

Table V Shares of COVID-19 journal publications that were preceded by a preprint, by signatory status (2020–2021)

Indicator	Non-signatory	RPO signatory	Funding signatory	Overall journal signatory	Non-Big 5 journal signatory
Share preceded by a preprint	5.3%	19.6%	19.4%	11.6%	12.2%
	[4.8,5.7]	[18.2,21.0]	[18.7,20.0]	[11.3,11.9]	[11.6,12.7]

Note: Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix

Findings on the dimension of prior preprinting for journal publications again show stark contrasts when broken down by signatory status. Non-signatory COVID-19 journal publications were preceded by preprints in only 5% of cases. For publications with signatory status through funding or institutional

[12] Fraser et al., The Evolving Role of Preprints in the Dissemination of COVID-19 Research and Their Impact on the Science Communication Landscape.

affiliation, this share jumps to almost 20%. Articles published with signatory journals were preceded by preprints in 12% of cases.

Controlled differential findings all converge on this indicator (Table VI). They show a clear increase in preprinting for signatory publications between 2018–19 and 2020–21. This increase is above an increase also recorded for non-signatory publications. The differential gain ranges between +5 percentage points and + 29 percentage points, depending on the DiD model considered. On this basis, it is possible to conclude that the Joint Statement likely contributed to increasing the share of journal articles preceded by a preprint.

Table VI Differential gains in shares of COVID-19 journal publications that were preceded by a preprint, by signatory status (2020–2021)

Indicator	Signatory definition	DiD with all articles	DiD with controlled authors	DiD with controlled, weighted authors
Share preceded by a preprint	fund OR journal	+4.6 p.p.	+19.0 p.p.	+12.8 p.p.
		[3.6,5.8]	[14.0,23.5]	[5.3,20.1]
	fund AND journal	+11.7 p.p.	+28.7 p.p.	+20.3 p.p.
		[10.3,13.3]	[16.9,40.0]	[0.5,39.7]

Note: Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix

2.5 Writing of data availability statements in journal publications

It should be noted that recall on the DAS datasets was somewhat low at 62%. Caution should therefore be used in analysing these findings, as they are likely to be under-estimates. The stability intervals in the tables do not capture the error rate associated with the formulation of the regex queries themselves and the associated under-estimations.

The descriptive statistics (Table VII) show that shares of COVID-19 publications where authors have written a data availability statement (DAS) ranged between 42% and 45%, both for signatory and non-signatory groups of journal articles.

The one exception comes from descriptive statistics for the group of signatory journals owned outside the major publishers, where DAS writing was identified in 76% of publications.

The DiD models do not provide the ability to trace back a differential gain in DAS writing to the Joint Statement, and they also provide inconclusive results as to whether DAS writing has increased with the pandemic, irrelevant of signatory status (Table VIII).

Table VII Shares of COVID-19 journal publications that included a data availability statement, by signatory status (2020–2021)

Indicator	Non-signatory	RPO signatory	Funding signatory	Overall journal signatory	Non-Big 5 journal signatory
Share with a DAS	43.5%	41.9%	43.8%	44.8%	76.2%
	[41.7,45.3]	[39.5,44.2]	[42.6,45.0]	[44.2,45.5]	[75.3,77.1]

Note: Publications for which records of fulltext content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for DAS writing measurements. With a recall of 62% on this indicator, the estimates presented here are likely to under-estimate the real frequency in DAS writing.

Source: Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

Table VIII Differential gains in shares of COVID-19 journal publications that included a data availability statement, by signatory status (2020–2021)

Indicator	Signatory definition	DiD with all articles	DiD with controlled authors	DiD with controlled, weighted authors
Share with a DAS	fund OR journal	+1.2 p.p.	+6.5 p.p.	-21.4 p.p.
		[-9.0,12.2]	[-17.9,31.6]	[-109.4,51.9]
	fund AND journal	-2.9 p.p.	-20.4 p.p.	N/C
		[-13.6,8.6]	[-44.6,0.5]	

Note: Publications for which records of fulltext content could not be obtained or were incomplete or missing were considered null and removed from the computation of this this indicator. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for DAS writing measurements.

Source: Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

2.6 Writing of data availability statements in preprints

The request for caution in interpreting DAS writing findings applies to findings on preprints just as well as to journal publications.

A share of close to 60% of signatory COVID-19 preprints contained a DAS, or 70% when considering a more controlled group of preprints where preprints across the four analytical groups share at least one author, to control for a number of subfield- or author-level factors (Table IX). These shares were roughly 12 percentage points above the corresponding scores for non-signatory COVID-19 preprints.

It was impossible to attribute signatory or non-signatory status to the vast majority of preprints (32,722 out of 37,742). Those preprints saw a prevalence of DAS writing measured at 44%.

Statistically robust controlled differential findings could not be obtained in the set of COVID-19 and comparable research preprints given low number of available observations (Table X). Instead, an additional source of descriptive analysis was obtained by disaggregating these findings by preprint server. It should be noted that medRxiv mandates DAS for its authors, and therefore its share of preprints with a DAS has been set to 100%. Signatory preprints in arXiv and SSRN tended to have a much higher share of their numbers containing a DAS relative to all preprints (16% to 9% and 36% to 10%, respectively). bioRxiv preprints saw a modest increase in DAS writing for those with signatory status compared to all preprints (31% to 28%).

Table IX Shares of COVID-19 preprints that included a data availability statement, by signatory status (2020–2021)

Indicator	Unknown status	Overall non-signatory	Overall signatory	Controlled non-signatory	Controlled signatory
Share with a DAS	44.0% [43.5,44.6]	40.5% [36.8,44.0]	58.6% [57.2,60.1]	52.9% [37.3,66.7]	69.5% [65.7,73.2]

Note: Preprints for which records of full-text content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals are shown in brackets below the bootstrapping point estimate. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for DAS writing measurements. medRxiv preprints are systematically considered to have a DAS.

Source: arXiv, bioRxiv, medRxiv, SSRN; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

Table X Shares of COVID-19 preprints that included a data availability statement, by signatory status and preprint server (2020–2021)

Preprint server	Overall preprints		Signatory preprints	
	COVID-19 volume	Share with DAS	COVID-19 volume	Share with DAS
arXiv	4,901	9.2% [8.4,10.0]	625	16.0% [13.1,19.0]
bioRxiv	4,312	28.4% [27.1,29.6]	1,535	31.1% [28.7,33.4]
medRxiv	14,415	100.0%	1,852	100.0%
SSRN	14,494	9.8% [9.3,10.3]	361	36.3% [31.3,41.3]

Note: Preprints for which records of full-text content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals are shown in brackets below the bootstrapping point estimate. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for DAS writing measurements.

Source: arXiv, bioRxiv, medRxiv, SSRN; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

2.7 Mentions of data depositions within journal publications

In considering data deposition findings reported here, it is crucial to note that those measurements are associated with large margins of error that could not be fully quantified. As a comparatively infrequent phenomenon, computation of recall on data deposition would have required manual parsing of several hundred publications. Manual parsing have been conducted on samples of 50 journal publications from each signatory breakdown. These manual estimates indicate that the text mining-based queries are likely to have under-estimated the shares of publications with mentions of data deposition within a DAS. While the share of missed publications with data deposition may be relatively small in the largest group of journal-signatory publications, it may be much larger for funding-signatory publications, RPO-signatory publications and non-signatory publications. We therefore advise great caution in using these numbers. The comparisons between signatory groups and the non-signatory group is positive in favor of the first with both methods, and margins of error do not overlap also in both cases. This conclusion appears robust. However, we advice against using these findings to estimate the preponderance of data deposition practices at the moment.

Descriptive findings (Table XI) show greatly varying degrees of data sharing by signatory status. Non-signatory COVID-19 publications made mention of repositories of interest (among those selected, see Appendix F) within a DAS section in 2% of cases. This contrasted with 9%–11% of publications with a DAS that also contained a repository mention within journal publications with signatory status due to funding support or institutional affiliation. For publications with a DAS and published with a signatory journal, the share of publications that also contained a repository mention was 6% overall, or 7% in signatory journals outside the major five publishers.

Table XI Shares of COVID-19 journal publications that included a mention of data deposition within a data availability statement, by signatory status (2020–2021)

Indicator	Non-signatory	RPO signatory	Funding signatory	Overall journal signatory	Non-Big 5 journal signatory
Among all publications, share that included both a DAS and data deposition					
Regex queries	2.0%	9.3%	10.6%	6.1%	7.3%
	[1.5,2.5]	[7.9,10.7]	[9.8,11.3]	[5.8,6.4]	[6.8,7.9]
Manual validation - (50 publications per sample)	6.0%	20.0%	30.0%	8.0%	
	[5.3,6.7]	[9.1,30.9]	[17.3,42.7]	[7.2,8.8]	

Note: Publications for which records of full-text content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for data deposition measurements.
Source: Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

In the controlled experiments with journal publications, many findings were inconclusive, and some findings could not be computed due to low numbers of available observations (Table XII). The few findings that are reliable provide contradictory results depending on whether they are based on overall available publications (with findings of differential losses for the Joint Statement) or on publications controlled for authors (with findings of strong differential gains associated with the Statement). On this basis, it is uncertain whether higher data deposition scores for signatory publications can be attributed to the Statement, to the pivot towards pandemic-era research more generally, or to pre-existing practices at signatory organisations.

Table XII Differential gains in shares of COVID-19 journal publications that included a mention of data deposition within a data availability statement, by signatory status (2020–2021)

Indicator	Signatory definition	DiD with all articles	DiD with controlled authors	DiD with controlled, weighted authors
Among all publications, share that included both a DAS and data deposition	fund OR journal	-2.2 p.p. [-4.5,1.3]	+4.6 p.p. [-1.9,9.9]	N/C
	fund AND journal	-1.5 p.p. [-4.8,2.5]	+12.2 p.p. [-0.4,23.3]	N/C

Note: Publications for which records of full-text content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals are shown in brackets below the bootstrapping point estimate. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for data deposition measurements.

Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

2.8 Mentions of data depositions within preprints

Please note that the important limitations mentioned above for the data deposition indicator apply to preprints as well as to journal publications.

Lower shares of preprints mentioned repositories in their DAS sections as compared to journal publications (Table XIII; compare to Table XII). This share was measured at 5% for signatory COVID-19 preprints, 3% for non-signatory preprints, and 2% for preprints of unknown signatory status. Signatory preprints recorded similar shares in both overall and controlled data sets compared to non-signatory preprints which had a slightly smaller share in the overall set. Differences across all groups should be taken lightly as the robust confidence intervals of measured scores were overlapping.

Reasons for lower frequency of data deposits in preprints as compared to journal articles might include the difficulty of finishing and polishing user-friendly data set releases in time for early publication as part of a preprint. This hypothesis was supported by anecdotal observations of preprint DAS containing formulations roughly to the effect that 'data sets **will** soon be made available on repository XYZ'. Alternatively, authors may have developed a habit to plan for open data releases in the final stages of manuscript preparation for formal publication; they may be caught unprepared by the suggestion to deposit data earlier in the publication cycle.

Table XIII Shares of COVID-19 preprints that included a mention of data deposition within a data availability statement, by signatory status (2020–2021)

Indicator	Unknown status	Overall non-signatory	Overall signatory	Controlled non-signatory	Controlled signatory
Among all preprints, share that included both DAS and data deposition	1.6% [1.5,1.7]	2.9% [1.9,4.3]	4.7% [4.1,5.4]	3.9% [0.0,9.8]	4.7% [3.1,6.6]

Note: Preprints for which records of full-text content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for data deposition measurements. Given the uncertainties found with the findings on data deposition in the journal publications, re-use and interpretation of these findings should be made with great caution.

Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

Statistically robust controlled differential findings could not be obtained in the set of COVID-19 and comparable research preprints given the low number of available observations. Considering server-level findings, signatory preprints were systematically associated with higher levels of data deposition than when all preprints were considered. medRxiv saw the highest shares of data deposition when preprints with a DAS and a data deposition mention are considered as a share of overall publications (Table XIV). However, medRxiv slips to the last position when considering only preprints with a data deposition mention as a share of preprints with a DAS (data not shown).

Table XIV Shares of COVID-19 preprints that included a mention of data deposition within a data availability statement, by signatory status and preprint server (2020–2021)

Preprint server	Overall preprints		Signatory preprints	
	COVID-19 volume	Share DAS+data dep	COVID-19 volume	Share DAS+data dep
arXiv	4,901	1.5% [1.2,1.9]	625	3.7% [2.2,5.3]
bioRxiv	4,312	2.8% [2.3,3.3]	1,535	4.1% [3.1,5.1]
medRxiv	14,415	3.2% [2.9,3.4]	1,852	5.5% [4.5,6.6]
SSRN	14,494	0.8% [0.7,0.9]	361	5.0% [2.8,7.5]

Note: Preprints for which records of fulltext content could not be obtained or were incomplete or missing were considered null and removed from the computation of this indicator. Stability intervals for sampling errors are shown in brackets below the bootstrapping point estimate. Note that non-sampling errors are not taken into account in these intervals and are not marginal for data deposition measurements. Given the uncertainties found with the findings on data deposition in the journal publications, re-use and interpretation of these findings should be made with great caution.

Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix

2.9 Policy-related uptake of COVID-19 journal publications

Descriptive findings (Table XV) for journal publications again show stronger performances on policy-related uptake for affiliation- and funding-based signatory papers (normalised indices of 21.1 and 22.1, where the average level in a given subfield and year is set at 1.0), followed by journal signatories (13.3). Non-signatory publications fared quite below the signatory groups on policy-related uptake (6.5) in the descriptive results.

Table XV Normalised shares of COVID-19 journal publications having received one or more citations in a policy-related document, by signatory status (2020–2021)

Indicator	Non-signatory	RPO signatory	Funding signatory	Overall journal signatory	Non-Big 5 journal signatory
Normalized policy-related uptake	6.5 [4.8,8.4]	21.1 [17.7,25.0]	22.1 [18.2,27.1]	13.3 [12.0,15.0]	11.7 [9.4,15.4]

Note: Publications with no metadata in PlumX were considered null and removed from the computation of this indicator. Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: Scopus, Overton and PlumX; processed by Science-Metrix

Controlled, differential findings showed that signatory COVID-19 journal publications benefitted from a great increase in policy-related uptake compared against comparable signatory publications from 2018–19 (Table XVI). Non-signatory publications also tended to record increases, but these were smaller than for signatory publications. The comparison of signatory COVID-19 increases to non-signatory COVID-19 increases would normally control for any 'COVID-19-specific effect' in intensified policy-related uptake. Note that not all DiD models have reached definitive statistical certainty on this indicator, although even for those that don't, results are still positive and point towards a much higher likelihood of differential gain rather than differential decrease. Accordingly, it is possible to conclude to a likely positive contribution of the Joint Statement towards policy-related uptake. Note, however, that we did not test whether signatory findings with a data sharing mention or a preprint, specifically, were more likely to see policy-related uptake.

Table XVI Differential gains in shares of COVID-19 journal publications that received policy attention, by signatory status (2020–2021)

Indicator	Signatory definition	DiD with all articles	DiD with controlled authors	DiD with controlled, weighted authors
Normalized policy-related uptake	fund OR journal	+6.4 [3.5,9.5]	+17.4 [-1.5,38.6]	+14.2 [-14.6,44.4]
	fund AND journal	+15.3 [10.6,21.7]	+16.1 [-10.7,44.9]	+9.56 [-12.5,37.2]

Note: Publications with no metadata in PlumX were considered null and removed from the computation of this indicator. Stability intervals are shown in brackets below the bootstrapping point estimate.

Source: Scopus, Overton and PlumX; processed by Science-Metrix

2.10 Policy-related uptake of COVID-19 preprints

Signatory COVID-19 preprints saw much higher policy-related uptake (22% with at least one policy-related citation) than non-signatory preprints (6%; see Table XVII). This lead for signatory preprints was reduced in the non-controlled preprint set (12% to 5%).

Preprints of unknown signatory status recorded a share of 9% of their numbers cited within the policy-related literature, placing this performance in the middle between the signatory and non-signatory preprints' scores. Of course, the group of preprints with unknown signatory status may well include signatory and non-signatory publications, for instance by authors with access to formal grant funding but who have omitted to write funding acknowledgements at the time of posting their preprint.

Table XVII Shares of COVID-19 preprints having received one or more citation(s) in a policy-related document, by signatory status (2020–2021)

Indicator	Unknown status	Overall non-signatory	Overall signatory	Controlled non-signatory	Controlled signatory
Policy-related uptake (non-normalised)	9.3% [8.9,9.7]	5.3% [3.6,7.2]	11.5% [10.5,12.6]	12.0% [4.0,22.0]	21.7% [18.5,25.1]

Note: Preprints with no metadata in PlumX were considered null and removed from the computation of this indicator. arXiv preprints were assessed separately and in entirety at Overton. Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: arXiv, bioRxiv, medRxiv, SSRN, Overton and PlumX; processed by Science-Metrix

As for the other preprint-based analyses here, statistically robust controlled differential findings could not be obtained in the set of COVID-19 and comparable research preprints given the low number of available observations.

The comparison of findings at the level of preprint servers (Table XVIII) shows that SSRN and medRxiv COVID-19 preprints have received the most policy-related attention, in both the overall group (15% and 12% of preprints with at least one mention in a policy-related document respectively) and the signatory subset (18% and 17). These results of higher policy-related uptake for SSRN and medRxiv are well-aligned with prior findings that policy-related attention and even online attention more generally towards research is often higher for disciplines in the social and medical sciences.[13] Remember that normalisation of the policy-related uptake indicator could not be performed in the publications set of preprints, which may have helped mitigate some of these field-related discrepancies, if distribution of preprints amongst disciplinary subfields differs between the signatory and non-signatory groups.

The medRxiv preprints posted the highest lead in policy-related uptake for signatory preprints against the overall group (a difference of almost 5 percentage points), although in the absence of robust DiD findings

[13] Pinheiro et al., A Large-Scale Validation of the Relationship between Cross-Disciplinary Research and Its Uptake in Policy-Related Documents, Using the Novel Overton Altmetrics Database; Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLOS ONE*, *10*(3), p. e0120495. doi:10.1371/journal.pone.0120495.

it is not possible to tell whether this difference is due to Statement-related changes or should be traced back to specific, pre-existing practices within signatory organisations.

Table XVIII Shares of COVID-19 preprints having received one or more citation(s) in a policy-related document, by signatory status and preprint server (2020-2021)

Preprint server	Overall preprints		Signatory preprints	
	COVID-19 volume	Policy-related uptake	COVID-19 volume	Policy-related uptake
arXiv	4,901	2.4% [2.0,2.9]	625	4.5% [3.0,6.1]
bioRxiv	4,312	6.6% [5.9,7.4]	1,535	8.2% [6.8,9.8]
medRxiv	14,415	11.8% [11.3,12.4]	1,852	16.5% [14.6,18.2]
SSRN	14,494	15.1% [13.8,16.4]	361	17.7% [10.4,25.0]

Note: Preprints with no metadata in PlumX were considered null and removed from the computation of this indicator. arXiv preprints were assessed separately and in entirety at Overton. Stability intervals are shown in brackets below the bootstrapping point estimate.
Source: arXiv, bioRxiv, medRxiv, SSRN, Overton and PlumX; processed by Science-Metrix

2.11 Assessing longevity of the Zika Statement outputs and outcomes

To try and get some early signal on the potential longevity of open access, data sharing and preprinting effects of the Joint Statement, parallels were drawn with the Zika Statement issued in 2016 by Wellcome and others. It was initially hoped that outputs and outcomes related to the Zika Statement would be comparable to those of the COVID-19 Joint Statement, and that an assessment of open access, data sharing and preprinting dimensions in Zika research over the period 2014–2020 could help cautiously predict what the longevity of the COVID-19 Statement's impacts and outcomes might be. As for COVID-19, journal publications and preprints were collected for a period prior to the Zika Statement to try and compute a baseline against which to measure effects of the Statement.

Figure 2 shows that there are few Zika publications with sufficient metadata to attribute signatory status for the years 2014 and 2015, a situation compounded by a very low original number of publications retrieved. Therefore, Zika publications will only be included in the analyses that follow starting with 2016, the year of the Joint Statement on Zika.

Also note that the practice of preprinting in Zika research, but also in comparable HVVD, has been restricted (Figure 3). Between 2016 and 2018 there were only 20 signatory Zika and HVVD preprints issued yearly (in each group), and five or fewer non-signatory Zika and HVVD preprints (each). These numbers are too low to allow for robust analyses of data sharing practices in preprints, and therefore these have been omitted here.

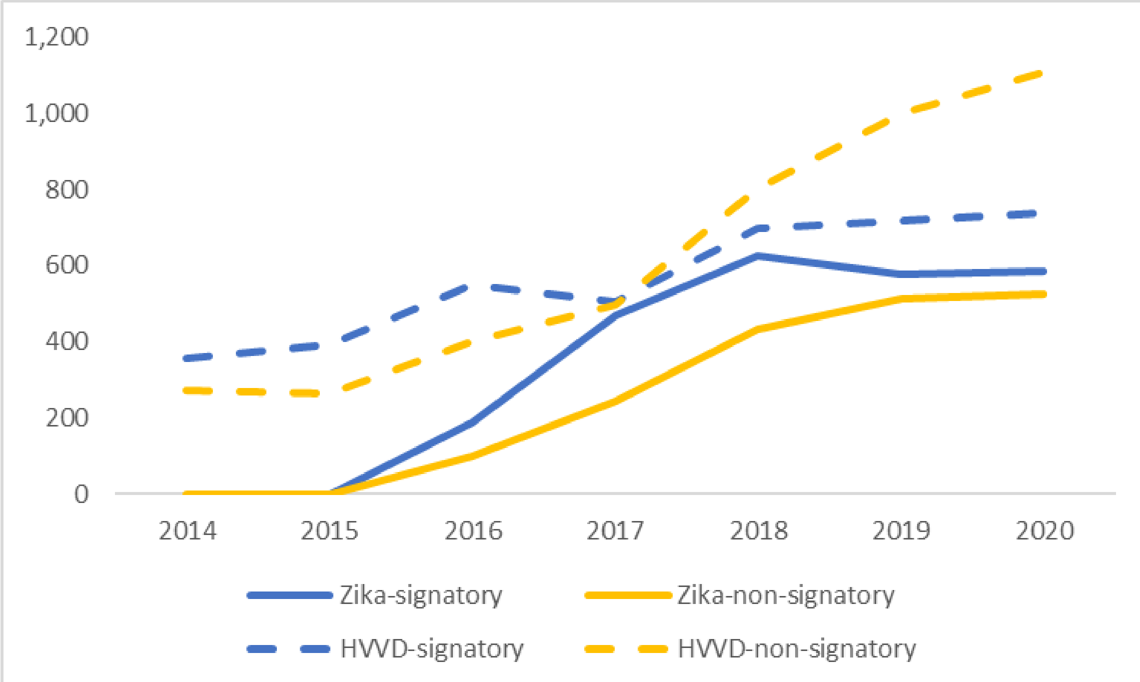

Figure 2 Journal publication volume, Zika and human vector-borne viral diseases, 2014–2020

Source: Scopus; processed by Science-Metrix

The Zika statement appeared to have mobilised a good portion of the potential (at the time) producers, funders, or publishers of future Zika research. Signatory publications made up 60% of Zika journal publications between 2016 and 2020. The share of Zika preprints falling in the signatory group was much higher, at 90%.

In terms of overall OA and free-to-read status of signatory Zika research (Figure 4), there seems to have been very little effect of the Zika Statement, with the figure remaining stable throughout the period for both signatory research strands. Both strands' scores range between 80% and 90% over the period.

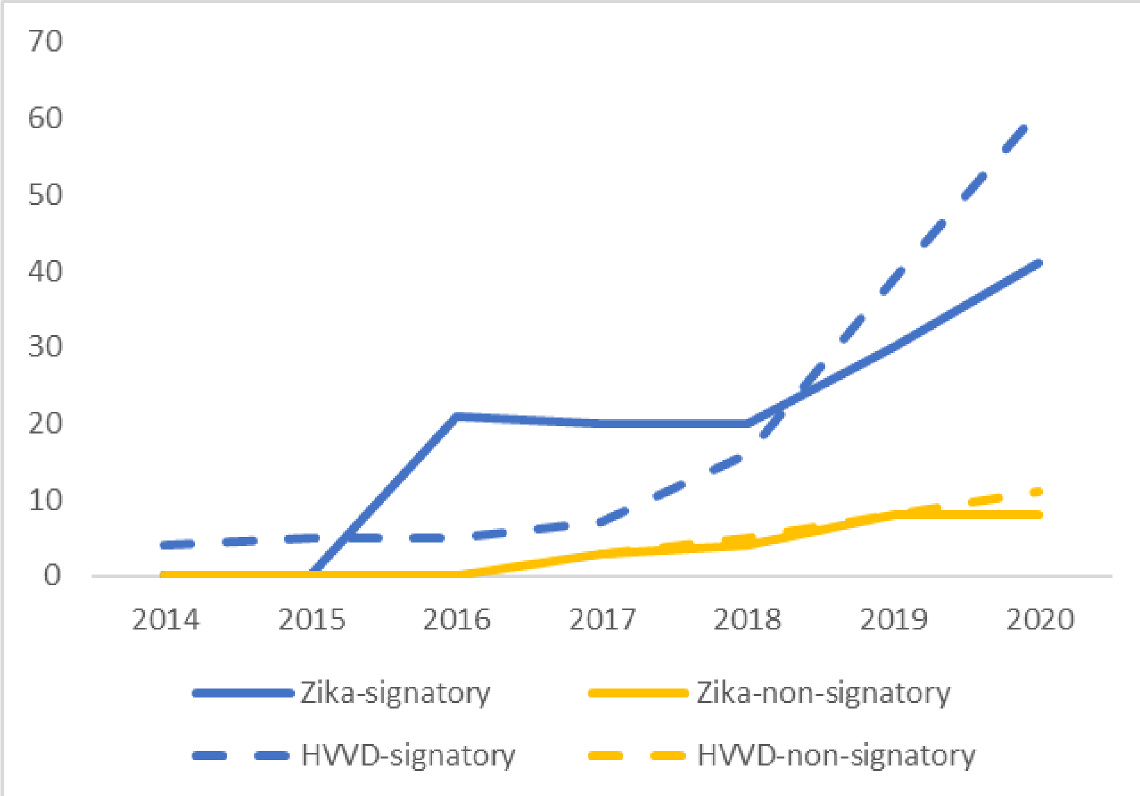

Figure 3 Preprints volume, Zika and human vector-borne viral diseases, 2014–2020

Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix

Signatory Zika research was available under a Gold OA licence in a proportion of 40% in 2016 (Figure 5). By 2020, that figure had risen to 57%. This increase roughly mirrored increases in the two non-signatory strands. Signatory HVVD research instead saw stable shares of journal publications available in gold OA increase from 47% to 60% over the same period. These observations indicate the Zika Statement was unlikely to have directly impacted the propensity to publish with gold OA journals. All analytical groups here have seen parallel, almost synchronized increases in share of journal publications available under a gold OA modality, pointing towards the effects of a secular trend.

Publisher-based free-to-read started (in 2016) at 24% for signatory Zika research, but this figure was halved by 2018 and settled at 10% in 2019 and 2020 (Figure 6). Non-signatory Zika research reached a share of 23% of articles available through publisher-based free-to-read status in 2017, but this share also decreased to 10% subsequently. The two non-signatory research strands' shares generally ranged between 5% and 12%.

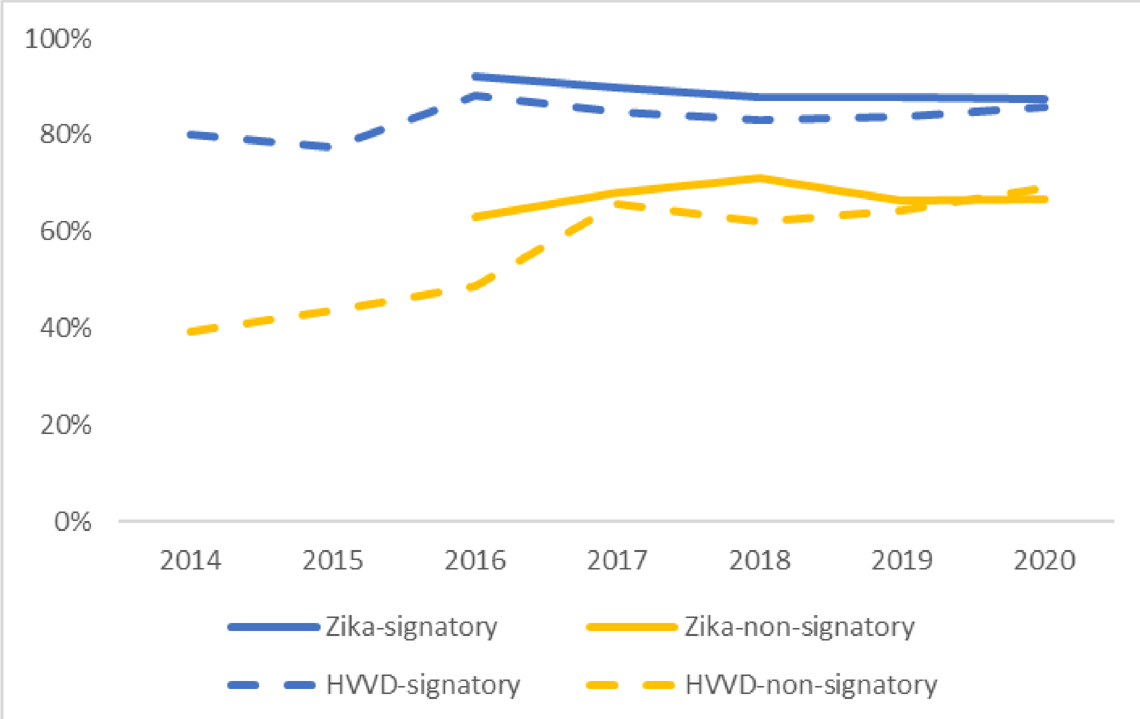

Figure 4 Overall share of journal publications available under an OA or free-to-read modality, Zika and human vector-borne viral diseases, 2014–2020

Source: Scopus, Unpaywall; processed by Science-Metrix

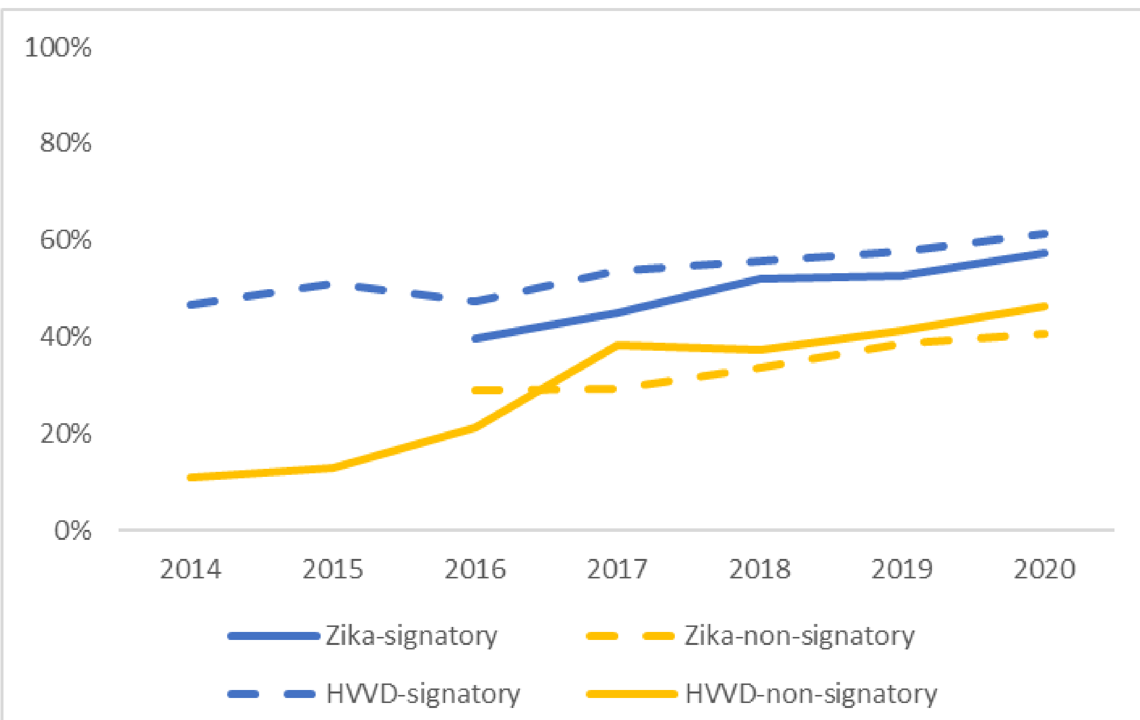

Figure 5 Share of journal publications available under a gold OA licence, Zika and human vector-borne viral diseases, 2014–2020

Source: Scopus, Unpaywall; processed by Science-Metrix

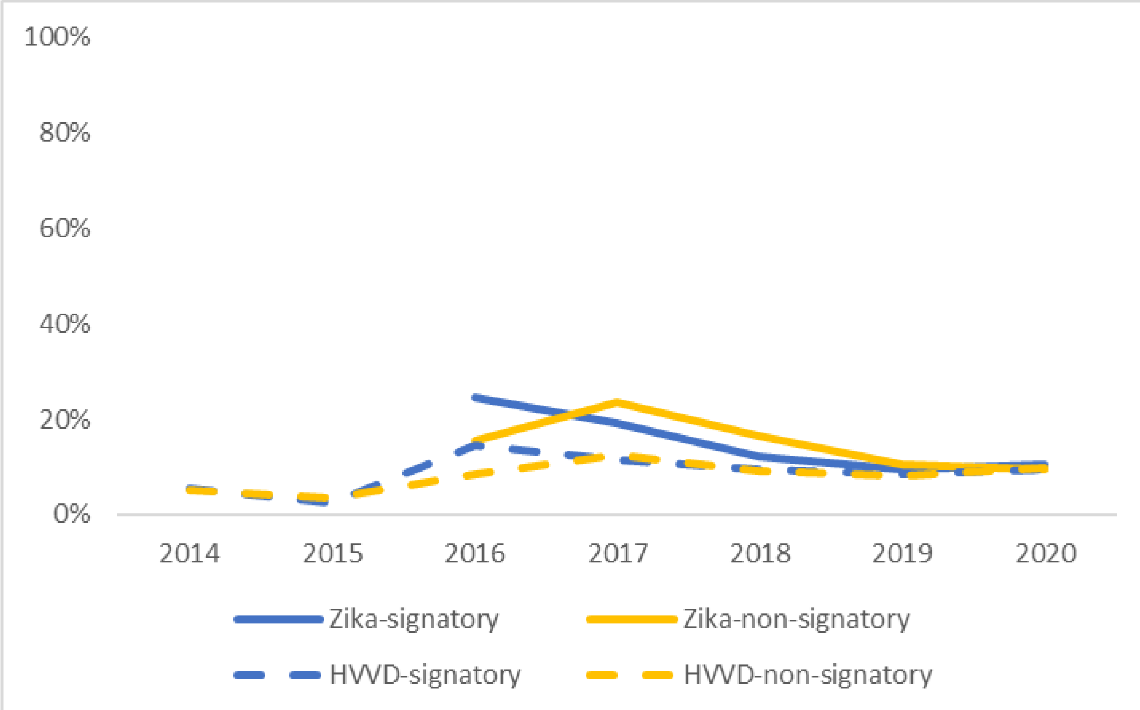

Figure 6 Share of journal publications available under a publisher-based free-to-read modality, Zika and human vector-borne viral diseases, 2014–2020

Source: Scopus, Unpaywall; processed by Science-Metrix

DAS writing in signatory Zika research saw a sharp initial increase from 2016 to 2017, somewhat plateauing afterwards (Figure 7). Signatory Zika research did not include DAS in any publications in 2016, followed by 19% in 2017. These scores fluctuated between 16% and 25% afterwards. Non-signatory Zika and HVVD research recorded DAS writing levels at roughly 10% or below through 2018 inclusively. By 2020, however, these numbers had gone up to more than 20%. These findings suggest a secular trend towards more DAS writing in the health sciences generally. Accordingly, the gap between non-signatory and signatory organisations is gradually decreasing owing to the now declining trend for signatories. This suggests a possible convergence of signatory and non-signatory research at a level close to signatory performance in 2016–2017. These findings suggest the possibility that increases in data sharing practices may not be durable for signatory organisations, although it is unclear why this affects DAS writing most out of all dimensions examined here.

In terms of mentions of data depositions made as part of DAS sections (Figure 8), signatory Zika research saw uneven increases from an initial null measurement in 2016. A share of 6% of signatory Zika journal publication included a DAS and data deposition in 2006, with this measurement fluctuating in the following years between 5% and 10%. Signatory HVVD research did not reach that propensity of data deposition, but did show yearly increases bringing the share to 7% in 2020.

The two non-signatory research trends steer quite close to their corresponding signatory groups and are characterised by a modest increases, much as was also the case for the two signatory trends. Combining observations from all four groups, it appears likely that the Zika Statement brought a temporary increase in data deposition culminating in 2017. The higher base level of data deposition thus reached in signatory

Zika research was then consolidated by both secular trends (from the similarity of trends with non-signatory research strands) but also possibly practices specific to signatory organisations. Alternatively, interventions aiming to strengthen data deposition specifically for Zika research might have spilled over into HVVD research as well.

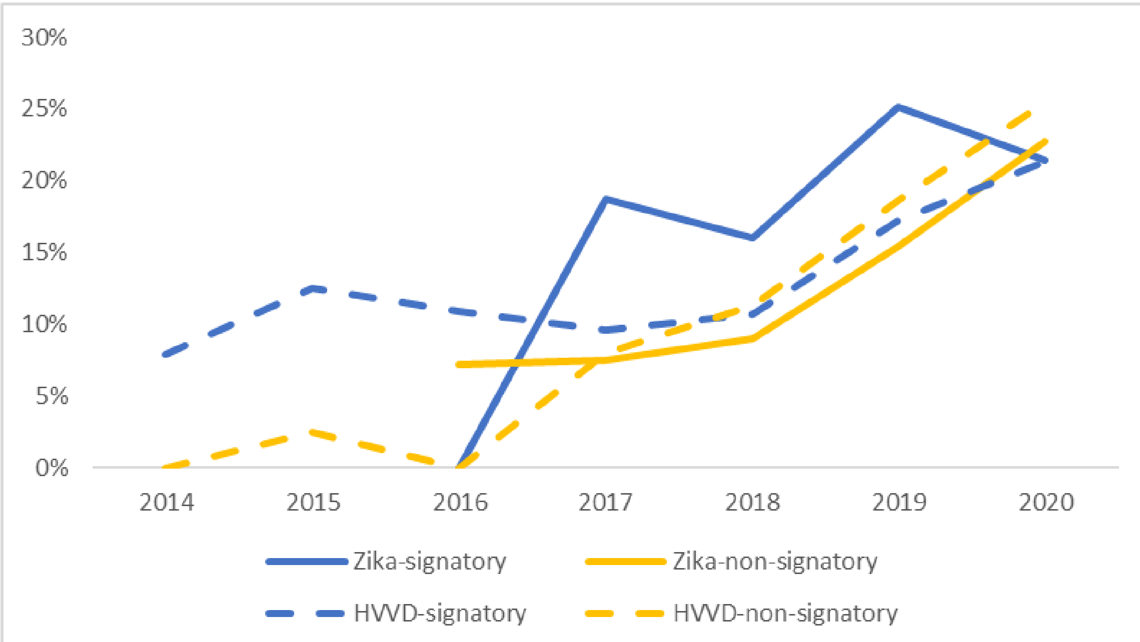

Figure 7 Share of journal publications containing a DAS section, Zika and human vector-borne viral diseases, 2014–2020

Source: Scopus; processed by Science-Metrix , Seyedamin Tabatabaei and Georgios Tsatsaronis

In terms of publications being preceded by a preprint (an indicator that is less affected by limitations due to low number of observations on preprints; Figure 9), there may have been an effect of the Zika Statement for a restricted period from 2016 to 2018. Roughly 5% of signatory Zika publications were preceded by a preprint over these three years, an unmatched level in any of the other analytical groups in those years.

In 2019 and 2020, preprinting increased from 5% to 6% and then 10% for signatory Zika publications. The corresponding increases in those years was greater for signatory HVVD research, going from 2% (2018) to 6% and then 9% (2020). These findings suggest that the Statement likely contributed an effect initially, but that this effect later combined with other, non-Statement-related practices to support preprinting at signatory organisations only. Those practices might, however, have been inspired by the Zika Statement. Non-signatory Zika research saw smaller increases in prior preprinting of journal publications over time, suggesting the existence of a secular trend towards increased preprinting, albeit of small magnitude.

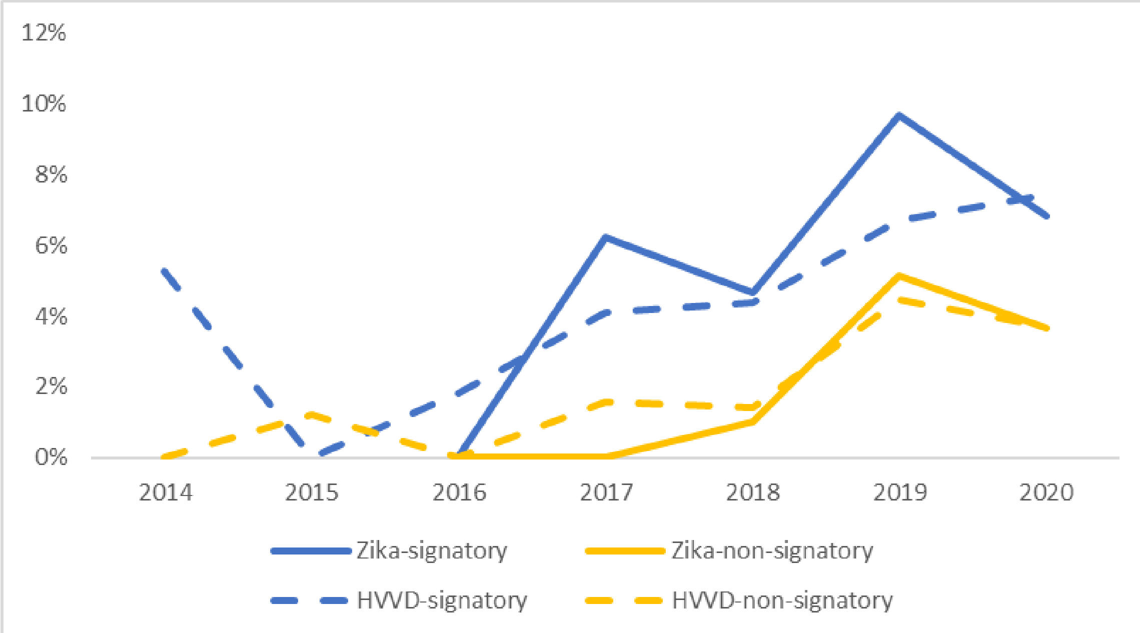

Figure 8 Share of journal publications containing a DAS section with a data deposition mention, Zika and human vector-borne viral diseases, 2014–2020

Source: Scopus; processed by Science-Metrix , Seyedamin Tabatabaei and Georgios Tsatsaronis

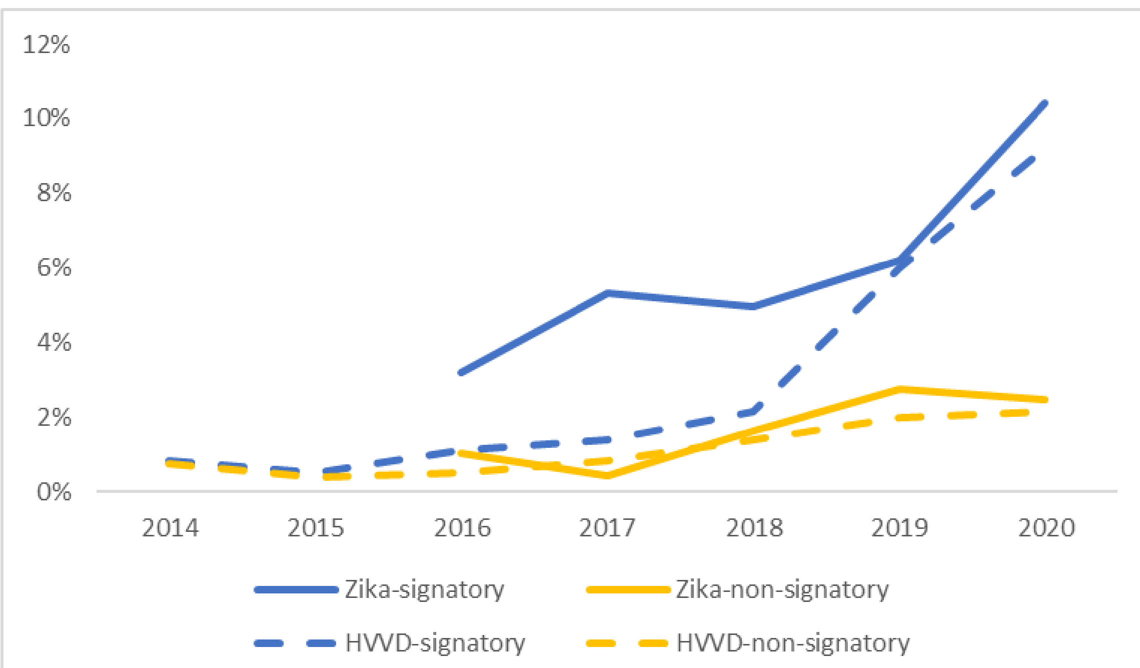

Figure 9 Share of journal publications preceded by a preprint, Zika and human vector-borne viral diseases, 2014–2020

Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix

To summarise findings on longevity of Zika Statement outputs and outcomes, generally, Zika research published by or supported by signatories to the 2016 Zika Statement did see higher levels of open access or free-to-read status, preprinting, and data sharing practice than non-signatory Zika research. That said, only for two dimensions were signatory Zika scores higher than comparable signatory research. These two dimensions are for prior preprinting (5% of signatory Zika journal articles being preceded by preprints in 2017, against less than 1% for non-signatory Zika research) and data deposition as part of a DAS (6% of Zika signatory articles in 2017, against 0% for Zika non-signatory research). Even in those two cases, the scores of comparable signatory research had reached the same levels as those of signatory Zika research by 2020.

In short, in many dimensions of the analysis there was no clear Zika Statement effect, for which the question of longevity becomes moot (OA and free-to-read status; DAS writing).

For the dimensions of preprinting and data deposition, a contribution of Zika Statement to higher signatory scores was likely for the years 2016 and 2017, but subsequent increases roughly mirrored those of the other analytical groups and are therefore likely to be less or not related to the Statement. Therefore, it appears the Zika Statement's effects were either of a small magnitude to start with; or where they reached greater magnitude, they did not persist after 2018.

Given our findings that the COVID-19 Statement signatories encompassed a much larger proportion of the corresponding research than the Zika Statement did, and also led to impacts of larger magnitudes than those of the Zika Statement, it now appears uncertain whether the second can inform us about the longevity of the first. Although the Zika Statement has had null or restricted effects and restricted longevity, this does not preclude different outcomes for the COVID-19 Joint Statement over the mid- or long-term.

3 Discussion and ramifications for the assessment of the Joint Statement theory of change

3.1 Summary of main findings

Generally, this study has been characterised by a high degree of conceptual complexity, with multiple layers of Joint Statement signatory status requiring a multiplication of specifications for DiD models and, as a result, of the findings themselves. Additionally, differences in metadata availability between preprints and journal publications brought in another layer of complexity, requiring distinct analyses for both types of research papers. Findings from across the analytical breakdowns and DiD models, where consistent, were interpreted as a sign of robustness.

Considering only descriptive statistics, groups of signatory publications systematically performed above non-signatory publications in terms of overall OA and free-to-read availability; research data deposition; prior preprinting for journal articles; and policy-related uptake. Considering these positive signals, the question remains: did the Joint Statement contribute to these performances above and beyond the broader changes brought by the pivot towards pandemic-related research?

Arguably the clearest positive findings from the bibliometric component of this study are that signatory journal publications were more often preceded by a preprint than non-signatory publications. This conclusion was supported by all DiD models. The Joint Statement thus appears very likely to have contributed to a differential gain in preprinting for signatory research, whereby **researchers would have posted signatory COVID-19 research on preprint servers more frequently than publications not supported by signatory funding or published with signatory journals.** Despite these incremental gains in preprinting for both Joint Statement signatories and COVID-19 research more generally, absolute measurements remain low, indicating that implementation of commitments to the Joint Statement have been partial at best.

Another strong signal came from policy-related uptake. Here again, all DiD models converged with positive results, although only about half of these findings were based on enough observations to achieve statistical robustness. It appears that **findings from signatory COVID-19 research were cited more frequently in the policy-related literature than non-signatory research.** Because the DiD models also take into account prior comparable research supported by the signatory and non-signatory organisations, this effect cannot be reduced solely to a factor of greater familiarity of governmental scientists or policymakers towards the group of signatory institutions.

The findings on the OA and free-to-read status of journal articles show clear and statistically robust increases in publisher-based free-to-read status for signatory publications relative to non-signatory publications. This finding comes with a relative decrease in Gold OA. These two trends create a complex picture for understanding overall OA and free-to-read status. The different DiD models used here do not converge into a single signal, and many findings are not statistically robust. The reason for this is that prior signatory research already recorded high shares of OA and free-to-read publications even before the pandemic, with negative or inconclusive DiDs simply capturing the fact that non-signatory research had more space available to increase its performanceon this dimension. Therefore, it cannot be definitively concluded that signatory publishers made journal COVID-19 publications immediately open access or free-to-read more frequently than non-signatory publishers. Nevertheless, it must be considered that the signatory organisations have succeeded in making 90% or more of signatory COVID-19 research available under an OA or free-to-read modality, and that non-signatory COVID-19 research has also seen increases in OA and free-to-read levels. **Therefore, it can be concluded that the COVID-19 publications have come closer to achieving full OA or free-to-read status than comparable prior research,** with this achievement driven by both signatory and non-signatory organisations.

Table XIX Summary of findings on differential gain in open science practices, by signatory status (2020–2021)

Indicator	**Signatory > non-signatory (descriptive, 20-2021)**	**Covid-19 > prior HVRD research (descriptive, 2020-21 vs 2018-19)**	**Joint Statement differential effect (DiD, 2020-21 vs 2018-19)**
Share of overall OA + free-to-read	jpubs: YES preprints: n/a	YES for signatory jpubs YES for non-signatory jpubs preprints: n/a	jpubs: inconclusive preprints: n/a
Share of gold OA	jpubs: YES for non-big5, NO otherwise preprints: n/a	NO for signatory jpubs YES for non-signatory jpubs preprints: n/a	jpubs: YES but negative preprints: n/a
Share of publisher-based free-to-read	jpubs: YES preprints: n/a	YES for signatory jpubs NO for non-signatory jpubs preprints: n/a	jpubs: YES preprints: n/a
DAS	jpubs: YES for non-big5, NO in overall group Preprints: YES	YES for signatory jpubs YES for non-signatory jpubs YES for signatory preprints YES for non-signatory preprints	jpubs: inconclusive preprints: inconclusive
DAS+ Data deposit	jpubs: YES preprints: inconclusive	NO for signatory jpubs YES for non-signatory jpubs preprints: inconclusive	jpubs: complex findings preprints: inconclusive
jpubs preceded by preprinting	jpubs: YES preprints: n/a	YES for signatory jpubs YES for non-signatory jpubs preprints : n/a	jpubs: YES preprints: inconclusive
Policy-related uptake	jpubs: YES preprints: YES	YES for signatory jpubs YES for non-signatory jpubs YES for signatory preprints YES for non-signatory preprints	jpubs: YES preprints: inconclusive

Note: 'Inconclusive findings' mean that bootstrapped confidence intervals are too broad to ascertain that the point estimates for DiD are statistically robust. 'Complex findings' means that results from different models are inconsistent and cannot be summarised. n/a: not applicable. "jpubs" : journal publications.
Source: arXiv, bioRxiv, medRxiv, Overton, PlumX, SSRN, Scopus,Unpaywall; processed by Science-Metrix

Finally, DAS- and data deposition-related findings are the study's most uncertain and complex findings. Considering only journal publications, signatory research did make mentions of data deposits in a greater proportion than non-signatory research. Yet DiD models provide contradictory findings, in some cases even recorded a differential decrease in data deposit for signatory publications. Given the low volume of observations available on data deposits (even more so in the smaller publication set of preprints), these findings need to be interpreted with extreme caution. In summary, **signatory journal publications did contain DAS with a mention of data deposit more frequently than non-signatory publications. However, it is still uncertain whether this difference can be attributed to the Joint Statement, to the pivot towards pandemic-era research, or to pre-existing practices at signatory organisations. Results for preprints are too uncertain to establish any robust conclusion in that set of publications.**

3.2 Summary of limitations

Limitations to the findings have been presented throughout this document and are summarized in the main report of the study. For additional prudence, they are recapitulated here again:

- The difference-in-differences design can control for many disciplinary, cultural, and author-level factors that might affect findings. It can also control for durable features or practices of signatory organisations and the researchers they support. Yet it cannot differentiate between the specific effects of the Joint Statement and new practices or initiatives by signatory organisations taken in response to the pandemic but independently of the Statement.
- In multiple metadata and full text processing steps such as isolating thematic sets of publications and preprints or retrieving mentions towards data sharing repositories, the study relied on manually curated lists of keywords and expressions. These queries are characterised by high precision (low share of false positives) but somewhat lower recall (somewhat higher share of false negatives, that is, imperfect coverage and representativity).
- Gaps in metadata and full text records in the datasets meant that signatory or non-signatory status could not be inferred for large portions of COVID-19 journal publications and preprints. In turn, this diminishes the representativity of our findings. To mitigate this issue, findings have been computed and provided separately for these publications and preprints of unknown signatory status. Availability of full text records, in a format workable for text mining to identify data availability statement and data deposition mentions within, was unevenly distributed among publishers, in part due to licensing issues.
- The combination of the two points above result in a good deal of uncertainty about the data deposition findings, particularly in respect to the frequency or intensity of data deposition in journal publications and preprints. The conclusion that signatory publications have comparatively higher frequency of data deposition against non-signatory publications is considered robust, however.
- Control groups for COVID-19 research, made up of "human viral respiratory diseases" (HVRD) journal publications and preprints, often contained significantly lower numbers of available observations, which has restricted the availability of robust difference-in-differences findings for some indicators.
- Coverage of preprint servers was limited to arXiv, bioRxiv, medrxiv, and SSRN only.

3.3 Recommendations for future quantitative monitoring of preprinting and data sharing

While the indicators presented in this study are simple from a conceptual and design point of view, the preparation and standardization of datasets to be used as input in these indicators' computation has required disproportionate data preparation efforts. To ease future quantitative monitoring assessments of preprinting and data deposition, all participant groups in the research-to-publication continuum can contribute to greater availability and quality of metadata on these practices.

Recommendations for publishing houses:

Monitoring of data deposition practices would be greatly enhanced if publishing houses standardized their use of DAS section, and mandated certain best practices from authors:

- Publishing houses should curate and make available DAS sections as standard metadata fields rather than as privileged components of publication full texts.
- DAS sections should be formatted so as to ease both machine reading but also manual qualitative assessments and validation of their contents.
- DAS sections should include a unique dataset identifier for each distinct mention of data deposition. Publishing houses should ease the input of these identifiers through pre-defined input fields and the provision of pre-defined options in addition to a freetext field. Classes of identifiers should be clearly identified for authors (i.e., (dataset-specific)-DOI, accession number, National Clinical Trial number, and so forth).
- DAS sections should be included in the manuscripts submitted to the peer review process. This way, peer reviewers can evaluate the quality and completeness of DAS sections. Doing so would also greatly enhance assessments of reproducibility or replication done as part of the peer-reviewing process. This recommendation poses a challenge for blinded review processes, which is in turn addressed in the next two bullet points.
- To allow blinded peer review of datasets (and their corresponding manuscript), publishing houses could develop anonymized online data repositories. Datasets deposited on these repositories could then be forwarded to established repositories once the peer-review process has been completed. To allow early deposit of datasets before the peer review process is activated, further coordination and joint publication processes should be achieved by both data repositories and publishing houses.

Recommendation for online data repositories and data sharing platforms

- To enable blinded review of data depositions and their corresponding manuscripts, online data repositories should coordinate with publishing houses and enable anonymized dataset deposition in parallel to early public dataset deposition, as described above.

Recommendation for publication authors

Arguably more so even than publishing houses, authors often have the last word on the content of preprints and of DAS sections. Therefore, we strongly encourage authors to consider:

- Providing comprehensive, accurate and close-to-final funding information for the preprints, just as they would do for a journal publication.

- Providing comprehensive, accurate and close-to-final metadata fields for their preprints.
- Providing comprehensive DAS sections that follow the guidelines provided above, even where they are not publisher-mandated. Authors should notably make more active use of unique dataset identifiers such as (dataset-specific)-DOIs, accession numbers, National Clinical Trial numbers, and so forth.

Recommendation for RPO and funders

Breakdowns of publications supported signatory funders or where one or more author(s) was affiliated to a signatory RPO recorded the highest scores on data deposition and preprinting. On this basis, it appears likely that RPO and funders support can realistically intervene on data sharing practices by researchers and publishers.

RPOs and funders should therefore continue to reflect on and evaluation how they can support affiliated or funded researchers' data sharing practices. RPOs and funders could notably provide templates for data availability statement sections and strongly advocate for the mention of unique dataset identifiers in these sections. RPOs and funders can also promote the mention of funding sources in preprints to affiliated or funded researchers. They can encourage these researchers to update preprint servers' self-reported information on final journal version(s) of their preprint(s).

3.4 Acknowledgements

Science-Metrix and the contributors to the bibliometric component of the study wish to express their gratitude to the members of the study Advisory Group (including representatives from the study sponsors) for stimulating discussions and thoughtful comments:

Hannah Hope, Sabah Adams (Wellcome Trust)

Rachel Bruce, Callum Day (UKRI)

Ashley Farley (Bill & Melinda Gates Foundation)

Alice Norton, University of Oxford

Elizabeth Gadd, University of Glasgow

Katharina Lauer, ELIXIR Europe

Ludo Waltman, Leiden University

Robert Terry, World Health Organisation

Appendix A – Completementary findings on OA, preprinting and data sharing in journal publications without funding acknowledgement metadata

To help assess the impact of the exclusion of more than 50% of COVID-19 journal publications exclusions the analyses due to lack of metadata on funding status of these publications, Table XX below provides measurements on the OA, preprinting and data sharing dimensions of interest within this aggregate.

Table XX OA, preprinting and data sharing measurements for COVID-19 journal publications with unknown signatory status (2020-2021)

Indicator	Measurement
Share of overall OA+free-to-read	83.4% [83.2,83.7]
Share of gold OA	34.9% [34.6,35.3]
Share of hybrid OA	7.6% [7.4,7.8]
Share of publisher-based free-to-read (bronze)	34.4% [34.0,34.7]
Share of repository-based free-to-read (green)	6.5% [6.4,6.7]
Share preceded by a preprint	4.5% [4.3,4.6]
Share with a DAS	39.6% [38.9,40.2]
Among all publications, share that included both a DAS and data deposition	2.2% [2.0,2.4]
Normalized policy-related uptake	6.7 [6.3,7.0]

Source: arXiv, bioRxiv, medRxiv, Overton, PlumX, Scopus, SSRN, Unpaywall; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

Appendix B - Complementary findings on OA, preprinting and data sharing in journal publications with at least one researcher based in a low income country or lower-middle income country

Late over the course of this study, a demand was formulated to try and assess the intensity of OA publishing, preprinting and data sharing in COVID-19 research with contributions from researchers situated on low income countries (LIC) or lower-middle income countries (LMIC; as defined by the World Bank). Table XXI provides evidence to support such an assessment.

Table XXI OA, preprinting and data sharing measurements for COVID-19 journal publications with one or more authors from a low income country or lower-middle income country, (2020-2021)

Indicator	Non-sig. LIC LMIC	Non-sig. LIC LMIC South-North	Signatory LIC LMIC	Signatory LIC LMIC South-North
Share of overall OA+free-to-read	85.0% [82.7,87.1]	87.2% [84.4,90.1]	91.5% [90.7,92.3]	93.2% [92.3,94.0]
Share of gold OA	64.7% [61.7,67.4]	65.8% [61.7,70.0]	37.6% [36.4,38.9]	42.9% [41.2,44.6]
Share of hybrid OA	6.7% [5.4,8.2]	7.1% [4.9,9.5]	6.3% [5.7,6.9]	8.8% [7.8,9.8]
Share of publisher-based free-to-read (bronze)	10.5% [8.7,12.3]	10.5% [7.9,13.0]	41.6% [40.3,42.8]	34.0% [32.4,35.6]
Share of repository-based free-to-read (green)	3.1% [2.2,4.2]	3.8% [2.2,5.5]	6.0% [5.4,6.7]	7.5% [6.5,8.3]
Share preceded by a preprint	5.0% [3.8,6.3]	5.8% [3.9,7.9]	8.2% [7.5,8.9]	8.8% [7.7,9.7]
Share with a DAS	45.7% [41.0,50.5]	42.7% [36.2,49.7]	35.9% [34.2,37.8]	44.2% [41.6,46.6]
Among all publications, share that included both a DAS and data deposition	0.9% [0.2,1.9]	1.5% [0.0,3.5]	3.6% [3.0,4.4]	4.3% [3.3,5.4]
Normalized policy-related uptake	3.4 [1.1,7.5]	4.6 [1.3,9.7]	8.4 [6.2,11.5]	9.3 [6.7,13.1]

Note: South-North: in addition to a LIC or LMIC-based author, the publication also includes one or more researchers from high income countries as defined by the World Bank. Bulgaria and China are exceptionally included in the list of high income countries for this indicator.
Source: arXiv, bioRxiv, medRxiv, Overton, PlumX, Scopus, SSRN, Unpaywall; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

Appendix C – Precision and recall of query-based datasets

Precision (share of true positives in a sample) and recall (share of retrieved records out of all relevant records) are the most common validation measurements employed in bibliometrics. As Table XXII shows, precision levels reached in the datasets employed are high, with precision levels of 95% or more being ideal. The precision scores for the HVRD dataset, and the combined DAS and data deposition dataset (92% in both cases), indicate that these findings should be re-used and interpreted with some caution. High recall figures are typically harder to achieve, and the usual threshold used by Science-Metrix in this test is a measurement of 70% of more to reach an adequate level of external validity.

Table XXII Precision and recall measurements for query -based datasets used in the study

Dataset under evaluation	Precision	Recall
Covid-19	100%	93%
HVRD	92%	76%
Zika	100%	92%
HVBVD	94%	88%
Signatory funders	99%	70%
Overall (signatory + non-signatory) funding-based attribution	N/C	78%
Signatory journals	N/C	97%
Signatory RPOs	100%	99%
Preprint-to-peer-reviewed-publication matching	94%	N/C
DAS identification and isolation	100%	62%
Data deposition within a DAS	92%	*

Note: N/C: not computed. *: see discussion below.
Source: arXiv, bioRxiv, medRxiv, SSRN, Scopus; processed by Science-Metrix, Seyedamin Tabatabaei and Georgios Tsatsaronis

Precision and recall figures are computed from manual assessment of samples of 50 randomly chosen journal publications. If initial measurements are below the 95% or 70% threshold, samples are extended to 100 further entries.

Note that computation of a recall measurement for the data deposition within a DAS dataset would have required manual parsing of several hundred publications, given the low proportions of data sharing found in journal publications. To compensate for this situation and allow a minimal level of triangulation, estimates based on manual assessment in a small sample were already provided above, along with associated margins of error.

Appendix D: Regex queries for assembling the disease-based thematic research publication sets

The following search terms were employed in the regex queries to delineate the COVID-19 research publication set:

- ncov.?19
- covid.?19
- sars.?cov.?2

- 2019.?ncov
- 2019 novel coronavirus
- novel coronavirus 2019
- coronavirus disease 2019

The following search terms were employed in the regex queries to delineate the HVRD research publication set:

- '(?i)\\basian flu\\b'
- '(?i)\\bcommon cold\\b'
- '(?i)\\bflu pandemic.*\\b'
- '(?i)\\bflu season\\b'
- '(?i)\\bflu shot.*\\b'
- '(?i)\\bflu vaccination\\b'
- '(?i)\\bflu vaccine.*\\b'
- '(?i)\\bh1n1\\b'
- '(?i)\\bh2n2\\b'
- '(?i)\\bh3n2\\b'
- '(?i)\\bh5n1\\b'
- '(?i)\\bhanta.?viridae\\b'
- '(?i)\\bhanta.?virus.*\\b'
- '(?i)\\bhong kong flu\\b'
- '(?i)\\bhuman corona.?virus.*\\b'
- '(?i)\\bhuman influenza\\b'
- '(?i)\\bhuman metapneumo.?virus.*\\b'
- '(?i)\\bhuman parainfluenza virus.*\\b'
- '(?i)\\bhuman rhino.?virus.*\\b'
- '(?i)\\binfluenza b\\b'
- '(?i)\\binfluenza epidemic.*\\b'
- '(?i)\\binfluenza forecast.*\\b'
- '(?i)\\binfluenza outbreak.*\\b'
- '(?i)\\binfluenza pandemic.*\\b'
- '(?i)\\binfluenza season\\b'
- '(?i)\\binfluenza vaccination\\b'
- '(?i)\\binfluenza vaccine.*\\b'
- '(?i)\\binfluenza virus b\\b'
- '(?i)\\bmers\\b'
- '(?i)\\bmiddle east respiratory syndrome\\b'
- '(?i)\\bpandemic flu\\b'
- '(?i)\\bpandemic influenza\\b'
- '(?i)\\brespiratory syncytial virus.*\\b'
- '(?i)\\brespiratory viral infection.*\\b'
- '(?i)\\brespiratory virus.*\\b'
- '(?i)\\bsars.?cov\\b'
- (?i)\\bseasonal flu\\b'
- '(?i)\\bseasonal influenza\\b'
- '(?i)\\bsevere acute respiratory syndrome\\b'
- '(?i)\\bspanish flu\\b'

- '(?i)\\bspanish influenza\\b'
- '(?i)\\bswine flu\\b'
- '(?i)\\bswine influenza\\b'
- '(?i)\\bviral bronchiolitis\\b'
- '(?i)\\bviral pneumonia\\b'
- '(?i)\\binfluenza\\b' AND ('(?i)\\bchild.*\\b' or '(?i)\\bhuman.*\\b' or '(?i)\\bpatient.*\\b' or '(?i)\\bvaccine.*\\b' or '(?i)\\bvaccination\\b' or '(?i)\\bepidemic.*\\b' or '(?i)\\bmorbidity\\b' or '(?i)\\bmortality\\b')

The following search terms were employed in the regex queries to delineate the Zika research publication set:

- '(?i)\\bzika\\b'
- '(?i)\\bzikv\\b'

The following search terms were employed in the regex queries to delineate the HVVD research publication set:

- '(?i)\\bbarmah forest virus.*\\b'
- '(?i)\\bbatai virus.*\\b'
- '(?i)\\bbourbon virus\\b'
- '(?i)\\bbunyamwera virus.*\\b'
- '(?i)\\bcalifornia encephalitis virus.*\\b'
- '(?i)\\bcalifornia serogroup virus.*\\b'
- '(?i)\\bcchfv\\b'
- '(?i)\\bchikungunya\\b'
- '(?i)\\bchikv\\b'
- '(?i)\\bchittoor virus.*\\b'
- '(?i)\\bdengue\\b'
- '(?i)\\bdenv\\b'
- '(?i)\\bfi.vre jaune\\b'
- '(?i)\\bheartland virus.*\\b'
- (?i)\\binkoo virus.*\\b'
- '(?i)\\bjamestown canyon virus.*\\b'
- '(?i)\\bjapanese encephalitis\\b'
- '(?i)\\bla.?crosse encephalitis\\b'
- '(?i)\\bla.?crosse virus.*\\b'
- '(?i)\\bmayaro fever\\b'
- '(?i)\\bmayaro virus.*\\b'
- '(?i)\\bmayv\\b'
- '(?i)\\bngari virus.*\\b'
- '(?i)\\bo.?nyong.?nyong virus.*\\b'
- '(?i)\\boropouche fever\\b'
- '(?i)\\boropouche virus.*\\b'
- '(?i)\\bphlebotomus fever\\b'
- '(?i)\\bpowassan virus.*\\b'
- '(?i)\\brift valley fever\\b'
- '(?i)\\bross river virus.*\\b'
- '(?i)\\brvfv\\b'
- '(?i)\\bsaint louis encephalitis\\b'

- '(?i)\\bsand.?fly fever\\b'
- '(?i)\\bsfts virus\\b'
- '(?i)\\bsftsv\\b'
- '(?i)\\bsindbis virus.*\\b'
- '(?i)\\bsnowshoe hare virus.*\\b'
- '(?i)\\bst. louis encephalitis\\b'
- '(?i)\\btahyna virus.*\\b'
- '(?i)\\busutu virus.*\\b'
- '(?i)\\bwest nile encephalitis\\b'
- '(?i)\\bwest nile virus.*\\b'
- '(?i)\\bwnv\\b'
- '(?i)\\byellow fever\\b'

Appendix E: Regex queries for identifying signatory funders

The following search terms were employed in funding acknowledgement sections to identify support by COVID-19 Joint Statement signatory funders:

African Academy of Sci

- '(?i)\\bAfrican[^a-z0-9]?Academy[^a-z0-9]?of[^a-z0-9]?Sci'
- '(?i)\\bacademie[^a-z0-9]?africaine[^a-z0-9]?des[^a-z0-9]?sci'
- '(?i)\\bAfrican[^a-z0-9]?Acad[^a-z0-9]{0,3}of[^a-z0-9]?Sci'
- '(?i)\\bAfrican[^a-z0-9]?Acad[^a-z0-9]{0,3}Sci'
- '(?i)deltas[^a-z0-9]?africa'
- '(?i)Alliance[^a-z0-9]?for[^a-z0-9]?Accelerating[^a-z0-9]?Excellence[^a-z0-9]?in[^a-z0-9]?Science'

UK Academy of Medical Sciences

- '(?i)\\bU[^a-z0-9]{0,1}K[^a-z0-9]{0,2}academy of Medical Sci'
- '(?i)\\bU[^a-z0-9]{0,1}K[^a-z0-9]{0,2}acad[^a-z0-9]{0,1}Med[^a-z0-9]{0,2}Sci'
- '(?i)\\bacademy of Medical Science[a-z]?[^a-z0-9]{0,2}U[^a-z0-9]{0,1}K\\b'
- '(?i)\\bacad[^a-z0-9]{0,1}Med[^a-z0-9]{0,2}Sci[a-z]?[^a-z0-9]{0,2}U[^a-z0-9]{0,1}K'
- '(?i)\\bU[^a-z0-9]{0,1}K[^a-z0-9]{0,2}academy of Medical Sci'
- '(?i)\\bU[^a-z0-9]{0,1}K[^a-z0-9]{0,2}acad[^a-z0-9]{0,1}Med[^a-z0-9]{0,2}Sci'
- '(?i)\\bacademy of Medical Science[a-z]?[^a-z0-9]{0,2}U[^a-z0-9]{0,1}K\\b'
- '(?i)\\bacad[^a-z0-9]{0,1}Med[^a-z0-9]{0,2}Sci[a-z]?[^a-z0-9]{0,2}U[^a-z0-9]{0,1}K'
- ('(?i)\\bClinician Scientist Fellowship\\b' AND '(?i)\\bacademy of Medical Sci')
- '(?i)\\bDaniel Turnberg Travel Fellow'
- '(?i)\\bGlobal Challenges Research Fund Networking Grant'
- '(?i)\\bNewton Advanced Fellow'
- '(?i)\\bNewton International Fellow'
- ('(?i)\\bSpringboard\\b' AND '(?i)\\bacademy of Medical Sci')
- ('(?i)\\bStarter Grant for Clinical Lecturer')
- ('(?i)\\bclinical lecturer starter grant' AND '(?i)\\bU[^a-z0-9]{0,1}K')

'Austrian Science Fund, FWF'

- '(?i)\\bAustrian[^a-z0-9]?Science[^a-z0-9]?Fund'
- '(?i)\\bF[^a-z0-9]{0,3}W[^a-z0-9]{0,3}F\\b'
- '(?i)\\bF[^a-z0-9]{0,3}F[^a-z0-9]{0,3}W[^a-z0-9]{0,3}F\\b'

- '(?i)\\bwissenschaftlichen[^a-z0-9]?Forsch'

'Gates Foundation, BMGF'

- '(?i)\\bGates[^a-z0-9]?FoundB'
- '(?i)\\bB[^a-z0-9]{0,1}M[^a-z0-9]{0,1}G[^a-z0-9]{0,1}F\\b'

Gulbenkian

- '(?i)\\bGulbenkian\\b'

Canada Foundation for Innovation

- '(?i)Canada[^a-z0-9]?Foundation[^a-z0-9]?for[^a-z0-9]?Inno'
- '(?i)\\bC[^a-z0-9]{0,1}F[^a-z0-9]{0,1}I\\b'
- '(?i)Canada[^a-z0-9]?Found[^a-z0-9]{0,3}Innov'
- '(?i)Canadian[^a-z0-9]?Foundation[^a-z0-9]?for[^a-z0-9]?Innovation'
- '(?i)Canadian[^a-z0-9]?Found[^a-z0-9]{0,3}Innov'

Canadian Institutes of Health Research

- '(?i)\\bC[^a-z0-9]{0,2}I[^a-z0-9]{0,2}H[^a-z0-9]{0,2}R\\b'
- '(?i)\\bCanadian[^a-z0-9]?Institutes[^a-z0-9]?of[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bCanada[^a-z0-9]?Institutes[^a-z0-9]?of[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bCanadian[^a-z0-9]?Institute[^a-z0-9]?of[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bCanada[^a-z0-9]?Institute[^a-z0-9]?of[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bCanadian[^a-z0-9]?Inst[^a-z0-9]?[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bCanada[^a-z0-9]?Inst[^a-z0-9]?[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bInstitut[a-z]?[^a-z0-9]?de[^a-z0-9]?recherche[^a-z0-9]?en[^a-z0-9]?sant.[^a-z0-9]?du[^a-z0-9]?Canada\\b'
- '(?i)\\bI[^a-z0-9]{0,1}R[^a-z0-9]{0,1}S[^a-z0-9]{0,1}C\\b'
- '(?i)\\bInstitut[a-z]?[^a-z0-9]?de[^a-z0-9]?recherche[^a-z0-9]?en[^a-z0-9]?sant.[^a-z0-9]?du[^a-z0-9]?Canada\\b'

Cancer Research UK, CRUK

- '(?i)\\bCancer[^a-z0-9]?Research[^a-z0-9]{0,3}U[^a-z0-9]{0,2}K\\b'
- '(?i)\\bCancer[^a-z0-9]?Res[^a-z0-9]{0,3}U[^a-z0-9]{0,2}K\\b'
- '(?i)\\bU[^a-z0-9]{0,2}K[^a-z0-9]{0,3}Cancer[^a-z0-9]?Res'
- '(?i)\\bC[^a-z0-9]{0,1}R[^a-z0-9]{0,1}U[^a-z0-9]?{0,1}K\\b'

US Centers for Disease Control and Prevention

- '(?i)\\bU[^a-z0-9]{0,1}S[^a-z0-9]{0,2}Center[a-z0-9]?[^a-z0-9]?(for|on|of)[^a-z0-9]?Disease[^a-z0-9]?Control[^a-z0-9]?(and|&)[^a-z0-9]?Prevention'
- '(?i)\\bUnited[^a-z0-9]?States[^a-z0-9]?Center[a-z0-9]?[^a-z0-9]?(for|on|of)[^a-z0-9]?Disease[^a-z0-9]?Control[^a-z0-9]?(and|&)[^a-z0-9]?Prevention'
- '(?i)\\bDisease[^a-z0-9]?Control[^a-z0-9]?(and|&)[^a-z0-9]?Prevention[^a-z0-9]U[^a-z0-9]{0,2}S'
- '(?i)\\bU[^a-z0-9]{0,1}S[^a-z0-9]{0,2}C[^a-z0-9]?D[^a-z0-9]?C'
- '(?i)\\bU[^a-z0-9]{0,1}S[^a-z0-9]{0,2}Center[a-z0-9]?[^a-z0-9]?for[^a-z0-9]?Disease[^a-z0-9]?Prevention[^a-z0-9]?(and|&)[^a-z0-9]?Contr'
- '(?i)\\bUnited[^a-z0-9]?States[^a-z0-9]?Center[a-z0-9]?[^a-z0-9]?for[^a-z0-9]?Disease[^a-z0-9]?Prevention[^a-z0-9]?(and|&)[^a-z0-9]?Contr'
- '(?i)\\bDisease[^a-z0-9]?Prevention[^a-z0-9]?(and|&)[^a-z0-9]?Control[^a-z0-9]U[^a-z0-9]{0,2}S'

CEPI Norway

- '(?i)\\bC[^a-z0-9]{0,1}E[^a-z0-9]{0,1}P[^a-z0-9]{0,1}I\\b' AND '(?i)norway')
- '(?i)Coalition[^a-z0-9]?(for|of)[^a-z0-9]?Epidemic[^a-z0-9]?Prepare'

Chinese Centre for Disease Control and Prevention

- '(?i)\\bChinese[^a-z0-9]?Centre[^a-z0-9]?(for|of|on)[^a-z0-9]?Disease[^a-z0-9]?Control[^a-z0-9]?(and|&)[^a-z0-9]?Pre'
- '(?i)\\bChina[^a-z0-9]?Centre[^a-z0-9]?(for|of|on)[^a-z0-9]?Disease[^a-z0-9]?Control[^a-z0-9]?(and|&)[^a-z0-9]?Pre'
- '(?i)\\bCentre[^a-z0-9]?(for|of|on)[^a-z0-9]?Disease[^a-z0-9]?Control[^a-z0-9]?(and|&)[^a-z0-9]?Prevention[^a-z0-9]{0,3}China'
- '(?i)\\bChina[^a-z0-9]?C[^a-z0-9]?D[^a-z0-9]?C'
- '(?i)\\bChinese[^a-z0-9]?C[^a-z0-9]?D[^a-z0-9]?C'

CIFAR

- '(?i)\\bCanadian[^a-z0-9]?Institute[a-z]?[^a-z0-9]?(for|of|on)[^a-z0-9]?Advanced[^a-z0-9]?Res'
- '(?i)\\bC[^a-z0-9]{0,1}I[^a-z0-9]{0,1}F[^a-z0-9]{0,1}A[^a-z0-9]{0,1}R\\b'
- '(?i)\\bInstitut Canadien de Recherches Avanc'
- '(?i)\\bI[^a-z0-9]{0,1}C[^a-z0-9]{0,1}R[^a-z0-9]{0,1}A\\b'
- '(?i)\\bCanada[^a-z0-9]?Institute[a-z]?[^a-z0-9]?(for|of|on)[^a-z0-9]?Advanced[^a-z0-9]?Res'
- '(?i)\\bCanada[^a-z0-9]?Inst[^a-z0-9]{0,3}Adv[^a-z0-9]{0,3}Res'
- '(?i)\\bCanadian[^a-z0-9]?Inst[^a-z0-9]{0,3}Adv[^a-z0-9]{0,3}Res'

CONCYTEC

- '(?i)\\bC[^a-z0-9]{0,2}O[^a-z0-9]{0,2}N[^a-z0-9]{0,2}C[^a-z0-9]{0,2}Y[^a-z0-9]{0,2}T[^a-z0-9]{0,2}E[^a-z0-9]{0,2}C'

Department of Biotechnology, India

- '(?i)\\(D[^a-z0-9]{0,1}B[^a-z0-9]{0,1}T\\)' AND ('(?i)\\bindia\\b' OR '(?i)\\bindian\\b' OR '(?i)\\bnew[^a-z0-9]?delhi\\b' OR '(?i)\\bMinistry[^a-z0-9]?of[^a-z0-9]?Science[^a-z0-9]?&[^a-z0-9]?Tech' OR '(?i)\\bMinistry[^a-z0-9]?of[^a-z0-9]?Science[^a-z0-9]?and[^a-z0-9]?Tech')
- '(?i)\\bMin[^a-z0-9]{0,3}Sci[^a-z0-9]{0,3}&[^a-z0-9]?Tech'
- '(?i)\\bMin[^a-z0-9]{0,3}Sci[^a-z0-9]{0,3}and[^a-z0-9]?Tech'

NWO

- '(?i)\\bDutch Research Council\\b'
- '(?i)\\bN[^a-z0-9]{0,1}W[^a-z0-9]{0,1}O\\b'
- ('(?i)Organisation for Scientific Research' OR '(?i)Organization for Scientific Research') AND '(?i)Netherlands'
- '(?i)Organisatie[^a-z0-9]?voor[^a-z0-9]?Wetenschappelijk[^a-z0-9]?Onderzoek'

European Molecular Biology Organization, EMBO

- '(?i)\\bE[^a-z0-9]{0,1}M[^a-z0-9]{0,1}B[^a-z0-9]{0,1}O[^a-z0-9]{0,1}\\b'
- '(?i)\\bEuropean[^a-z0-9]?Molecular[^a-z0-9]?Biology[^a-z0-9]?org'
- '(?i)\\bEuropean[^a-z0-9]?Molecular[^a-z0-9]?Biological[^a-z0-9]?org'
- '(?i)\\bE[^a-z0-9]{0,1}U[^a-z0-9]{0,2}Molecular[^a-z0-9]?Biology[^a-z0-9]?org'
- '(?i)\\bE[^a-z0-9]{0,1}U[^a-z0-9]{0,2}Molecular[^a-z0-9]?Biological[^a-z0-9]?org'

European Commission

- '(?i)\\bEuropean[^a-z0-9]?Commission\\b'
- '(?i)\\bE[^a-z0-9]{0,2}U[^a-z0-9]{0,2}Commission\\b'
- '(?i)\\bCommission Europ'
- '(?i)\\bEuropean Research Council\\b'
- '(?i)\\bE[^a-z0-9]?R[^a-z0-9]?C\\b'
- '(?i)\\bUnion's Horizon\\b'
- '(?i)\\bUnion Horizon\\b'
- '(?i)\\bE[^a-z0-9]{0,1}U[^a-z0-9]{0,2}Horizon\\b'
- '(?i)\\bHorizon[^a-z0-9]{0,2}2020\\b'
- '(?i)\\bFramework Programme\\b'
- '(?i)\\bF[^a-z0-9]{0,1}P.?1\\b|\\bF[^a-z0-9]{0,1}P.?2\\b|\\bF[^a-z0-9]{0,1}P.?3\\b|\\bF[^a-z0-9]{0,1}P.?4\\b|\\bF[^a-z0-9]{0,1}P.?5\\b|\\bF[^a-z0-9]{0,1}P.?6\\b|\\bF[^a-z0-9]{0,1}P.?7\\b|\\bF[^a-z0-9]{0,1}P.?8\\b|\\bF[^a-z0-9]{0,1}P.?9\\b'

Fondation Merieux

- '(?i)\\bFondation[^a-z0-9]?M.rieux' '
- '(?i)\\bFondation[^a-z0-9]?Christophe[^a-z0-9]?M.rieux'
- '(?i)\\bFondation[^a-z0-9]?Rodolphe[^a-z0-9]?M.rieux'
- '(?i)\\bFondation[^a-z0-9]?marcel[^a-z0-9]?M.rieux'
- '(?i)\\bRodolphe[^a-z0-9]?M.rieux[^a-z0-9]?fondation'
- '(?i)\\bRodolphe[^a-z0-9]?M.rieux[^a-z0-9]?foundation'
- '(?i)\\bmarcel[^a-z0-9]?M.rieux[^a-z0-9]?fondation'
- '(?i)\\bmarcel[^a-z0-9]?M.rieux[^a-z0-9]?foundation'
- '(?i)\\bM.rieux[^a-z0-9]?fondation'
- '(?i)\\bM.rieux[^a-z0-9]?foundation'

Fonds de recherche du Québec

- '(?i)\\bFond[a-z]?[^a-z0-9]?de[^a-z0-9]?recherche[^a-z0-9]?du[^a-z0-9]?Qu.bec\\b'
- '(?i)\\bF[^a-z0-9]{0,1}R[^a-z0-9]{0,1}Q\\b'

FDA

- '(?i)Drug[^a-z0-9]?Administra'
- '(?i)\\bF[^a-z0-9]?D[^a-z0-9]?A\\b' AND ('(?i)\\btobacco\\b' OR '(?i)\\bDrug[a-z?]\\b' OR '(?i)\\bfood\\b' OR '(?i)\\bC[^a-z0-9]?B[^a-z0-9]?E[^a-z0-9]?R\\b' OR '(?i)\\bC[^a-z0-9]?D[^a-z0-9]?E[^a-z0-9]?R\\b')
- ('(?i)\\bF[^a-z0-9]?D[^a-z0-9]?A\\b' AND ('(?i)\\bU[^a-z0-9]?S\\b' OR '(?i)\\bU[^a-z0-9]?S[^a-z0-9]?A\\b' OR '(?i)\\bUnited[^a-z0-9]?state')
- '(?i)\\bF[^a-z0-9]?D[^a-z0-9]?A[^a-z0-9]?grant\\b'
- '(?i)\\bF[^a-z0-9]?D[^a-z0-9]?A[^a-z0-9]?support\\b'
- '(?i)\\bfunded[^a-z0-9]?by[^a-z0-9]?the[^a-z0-9]?F[^a-z0-9]?D[^a-z0-9]?A\\b'
- '(?i)\\bsupported[^a-z0-9]?by[^a-z0-9]?the[^a-z0-9]?F[^a-z0-9]?D[^a-z0-9]?A\\b'

US Department of Health and Human Service

- '(?i)Department[^a-z0-9]?(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Human[^a-z0-9]?Service'
- '(?i)Dep[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?Health(and|&)[^a-z0-9]?Human[^a-z0-9]?Service'

- '(?i)Dpt[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Human[^a-z0-9]?Service'
- '(?i)Department[^a-z0-9]?(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Education[^a-z0-9]?(and|&)[^a-z0-9]?Welfare'
- '(?i)Department[^a-z0-9]?(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Welfare'
- '(?i)Department[^a-z0-9]?(of|for|on)[^a-z0-9]?H[^a-z0-9]?E[^a-z0-9]?W'
- '(?i)Department[^a-z0-9]?(of|for|on)[^a-z0-9]?H[^a-z0-9]?H[^a-z0-9]?S'
- '(?i)Dep[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Education[^a-z0-9]?(and|&)[^a-z0-9]?Welfare'
- '(?i)Dep[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Welfare'
- '(?i)Dep[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?H[^a-z0-9]?E[^a-z0-9]?W'
- '(?i)Dep[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?H[^a-z0-9]?H[^a-z0-9]?S'
- '(?i)Dpt[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Education[^a-z0-9]?(and|&)[^a-z0-9]?Welfare'
- '(?i)Dpt[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Welfare'
- '(?i)Dpt[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?H[^a-z0-9]?E[^a-z0-9]?W'
- '(?i)Dpt[^a-z0-9]{0,2}(of|for|on)[^a-z0-9]?H[^a-z0-9]?H[^a-z0-9]?S'
- '(?i)Center[^a-z0-9]?(for|of|on)[^a-z0-9]?Biologic[a-z]?[^a-z0-9]?Evaluation[^a-z0-9]?(and|&)[^a-z0-9]?Research'
- '(?i)Center[^a-z0-9]?(for|of|on)[^a-z0-9]?Device[a-z]?[^a-z0-9]?(and|&)[^a-z0-9]?Radiological Hea'
- '(?i)Center[^a-z0-9]?(for|of|on)[^a-z0-9]?Drug[^a-z0-9]?Evaluation[^a-z0-9]?(and|&)[^a-z0-9]?Research'
- '(?i)Center[^a-z0-9]?(for|of|on)[^a-z0-9]?Food[^a-z0-9]?Safety[^a-z0-9]?(and|&)[^a-z0-9]?Applied Nutrition'
- '(?i)\\bC[^a-z0-9]?F[^a-z0-9]?S[^a-z0-9]?A[^a-z0-9]?N\\b'
- '(?i)Center[^a-z0-9]?(for|of|on)[^a-z0-9]?Tobacco[^a-z0-9]?Produc'
- '(?i)Center[^a-z0-9]?(for|of|on)[^a-z0-9]?Veterinary[^a-z0-9]?Medicine'

French National Research Agency

- '(?i)French[^a-z0-9]?National[^a-z0-9]?Research[^a-z0-9]?Agency\\b'
- '(?i)French[^a-z0-9]?Research[^a-z0-9]?Agency\\b'
- '(?i)French[^a-z0-9]?National[^a-z0-9]?funding[^a-z0-9]?Agency\\b'
- '(?i)French[^a-z0-9]?National[^a-z0-9]?Agency[^a-z0-9]?(for|on|of)[^a-z0-9]?Res'
- '(?i)Agence[^a-z0-9]?National[a-z]?[^a-z0-9]?de[^a-z0-9]?la[^a-z0-9]?Recherche\\b'
- '(?i)\\bA[^a-z0-9]?N[^a-z0-9]?R\\b'

Genome Québec

- '(?i)G.nome[^a-z0-9]?Qu.bec'

Genome Canada'

- '(?i)G.nome[^a-z0-9]?Canada'

Health Research Board Ireland

'(?i)Health[^a-z0-9]?Research[^a-z0-9]?Board' AND ('(?i)Ireland' OR '(?i)Irish')

INSERM

- '(?i)\\bI[^a-z0-9]{0,1}N[^a-z0-9]{0,1}S[^a-z0-9]{0,1}E[^a-z0-9]{0,1}R[^a-z0-9]{0,1}M\\b'

- '(?i)French[^a-z0-9]?National[^a-z0-9]?Institute[^a-z0-9]?(for|of)[^a-z0-9]?Health[^a-z0-9]?(and|&)[^a-z0-9]?Medical[^a-z0-9]?Research'
- '(?i)Institut[^a-z0-9]?national[^a-z0-9]?de[^a-z0-9]?la[^a-z0-9]?sant.[^a-z0-9]?et[^a-z0-9]?de[^a-z0-9]?la[^a-z0-9]?recherche[^a-z0-9]?m.dical'

Institute of Tropical Medicine

- '(?i)\\bInstitute[a-z]?[^a-z0-9]?(of|for|on)[^a-z0-9]?Tropical[^a-z0-9]?Med' AND ('(?i)\\bBelgium\\b' OR '(?i)\\bAntwer')
- '(?i)\\bInstitut[^a-z0-9]?de[^a-z0-9]?m.decine[^a-z0-9]?Tropic' AND ('(?i)\\bBelgium\\b' OR '(?i)\\bAntwer')

Japan Agency for Medical Research and Development

- '(?i)\\bA[^a-z0-9]{0,2}M[^a-z0-9]{0,2}E[^a-z0-9]{0,2}D\\b'
- '(?i)Agency[^a-z0-9]?(for|on|of)[^a-z0-9]?Medical[^a-z0-9]?Res' AND '(?i)Japan'

Luxembourg National Research Fund

- '(?i)National[^a-z0-9]?Research[^a-z0-9]?Fund' AND '(?i)Luxembourg'
- '(?i)Fond[a-z]?[^a-z0-9]?National[^a-z0-9]?de[^a-z0-9]?la[^a-z0-9]?Recherche' AND '(?i)Luxembourg'
- '(?i)\\bF[^a-z0-9]{0,1}N[^a-z0-9]{0,1}R\\b' AND '(?i)Luxembourg'

Doctors without borders

- '(?i)\\bM.decins[^a-z0-9]?Sans[^a-z0-9]?Fronti'
- '(?i)\\bM.decin[^a-z0-9]?Sans[^a-z0-9]?Fronti'
- '(?i)\\bDoctors Without Borders\\b'

Michael Smith Foundation for Health Research

- '(?i)\\bMichael[^a-z0-9]?Smith[^a-z0-9]?Found'
- '(?i)\\bM[^a-z0-9]{0,1}S[^a-z0-9]{0,1}F[^a-z0-9]{0,1}H[^a-z0-9]{0,1}R\\b'

National Academy of Medicine, USA

- ('(?i)\\bNational[^a-z0-9]?Academy[^a-z0-9]?(of|)[^a-z0-9]?Med' OR '(?i)\\bNat[^a-z0-9]{0,2}Acad[^a-z0-9]{0,2}(of|)[^a-z0-9]?Med' OR '\\bNAM\\b' OR '(?i)\\bInstitute[^a-z0-9]?of[^a-z0-9]?Medicine\\b' OR '(?i)\\bInst[^a-z0-9]{0,2}?of[^a-z0-9]?Med') OR '\\bIoM\\b')

AND
('(?i)\\bU[^a-z0-9]{0,2}S\\b' OR '(?i)\\bUnited[^a-z0-9]?States\\b' OR '(?i)\\bamerican\\b')

National Institute for Health Research, UK

- '(?i)National[^a-z0-9]?Institute[^a-z0-9]?(for|on|of)[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)Nat[^a-z0-9]{0,2}Inst[^a-z0-9]{0,2}(for|on|of|)[^a-z0-9]?Health[^a-z0-9]?Res'
- '(?i)\\bU[^a-z0-9]{0,2}C[^a-z0-9]{0,2}L\\b' AND '(?i)Bio[^a-z0-9]?medical'

National Institute for Infectious Diseases Lazzaro Spallanzani

- '(?i)\\bLazzaro[^a-z0-9]Spallan\\B'

US National Institutes of Health

- '(?i)\\bNational[^a-z0-9]Institute[a-z]?[^a-z0-9]of[^a-z0-9]Hea[a-z]{0,3}\\b'
- '(?i)\\bN[^a-z0-9]{0,1}I[^a-z0-9]{0,1}H\\b'

- ('(?i)\\bNational[^a-z0-9]Cancer[^a-z0-9]Inst' AND ('(?i)\\bU[^a-z0-9]{0,2}S\\b' OR '(?i)\\bUnited States\\b'))
- ('(?i)\\bNational[^a-z0-9]Cancer[^a-z0-9]Inst' AND '(?i)\\bN[^a-z0-9]{0,1}C[^a-z0-9]{0,1}I\\b')
- '(?i)\\bN[^a-z0-9]{0,1}C[^a-z0-9]{0,1}I[^a-z0-9]{0,2}grant'
- '(?i)\\b(supported|funded) by N[^a-z0-9]{0,1}C[^a-z0-9]{0,1}I\\b'
- '(?i)\\bN[^a-z0-9]{0,1}C[^a-z0-9]{0,1}I[^a-z0-9]{0,2}U[^a-z0-9]{0,2}S'
- '(?i)\\bCenter[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Information[^a-z0-9]{0,2}Tech'
- '(?i)\\bCenter[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Scientific[^a-z0-9]{0,2}Review\\b'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Child[^a-z0-9]{0,2}Health[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Human Dev'
- '(?i)\\bFogarty[^a-z0-9]{0,2}International[^a-z0-9]{0,2}Cent'
- '(?i)\\bNational[^a-z0-9]{0,2}Center[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Advancing[^a-z0-9]{0,2}Translational[^a-z0-9]{0,2}Scien'
- '(?i)\\bNational[^a-z0-9]{0,2}Center[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Complementary[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Integrative[^a-z0-9]{0,2}Health\\b'
- '(?i)\\bNational[^a-z0-9]{0,2}Eye[^a-z0-9]{0,2}Inst'
- '(?i)\\bNational[^a-z0-9]{0,2}Heart[^a-z0-9]{0,2}Lung[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Blood[^a-z0-9]{0,2}Inst'
- '(?i)\\bNational[^a-z0-9]{0,2}Human[^a-z0-9]{0,2}Genome[^a-z0-9]{0,2}Research[^a-z0-9]{0,2}Inst' rads_2021_UK_WT_covidstatement.WT_signatory_inst_paper_ids_regex
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Allergy[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Infectious[^a-z0-9]{0,2}Disease'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Arthritis[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Muscu'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Bio[a-z0-9]?medical[^a-z0-9]{0,2}Imaging[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Bioengin'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Dental[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Cranio'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Diabete[a-z0-9]?[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Digestive'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Environmental[^a-z0-9]{0,2}Health[^a-z0-9]{0,2}Sci'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}General[^a-z0-9]{0,2}Medical[^a-z0-9]{0,2}Sci'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Mental[^a-z0-9]{0,2}Health\\b'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)f[^a-z0-9]{0,2}Neurological[^a-z0-9]{0,2}Disorder[a-z0-9]?[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Stroke\\b'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Nursing[^a-z0-9]{0,2}Res'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Aging\\b'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Alcohol[^a-z0-9]{0,2}Abuse[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Alco'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Deafnes[a-z0-9]?[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Other[^a-z0-9]{0,2}Com'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Drug[^a-z0-9]{0,2}Abuse\\b'
- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Minority[^a-z0-9]{0,2}Health[^a-z0-9]{0,2}(and|&)[^a-z0-9]{0,2}Health[^a-z0-9]{0,2}Dispa'

- '(?i)\\bNational[^a-z0-9]{0,2}Institute[a-z]?[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Health[^a-z0-9]{0,2}Clinical[^a-z0-9]{0,2}Cent'
- '(?i)\\bOffice[^a-z0-9]{0,2}of[^a-z0-9]{0,2}the[^a-z0-9]{0,2}Director\\b'
- ('(?i)\\bNational[^a-z0-9]{0,2}Library[^a-z0-9]{0,2}(for|of|on)[^a-z0-9]{0,2}Med' AND ('(?i)\\bU[^a-z0-9]{0,2}S\\b' OR '(?i)\\bUnited States\\b') OR '(?i)\\bN[^a-z0-9]?L[^a-z0-9]?M\\b')
- '(?i)\\bN[^a-z0-9]?C[^a-z0-9]?A[^a-z0-9]?T[^a-z0-9]?S\\b'
- '(?i)\\bN[^a-z0-9]?C[^a-z0-9]?C[^a-z0-9]?I[^a-z0-9]?H\\b'
- '(?i)\\bN[^a-z0-9]?H[^a-z0-9]?G[^a-z0-9]?R[^a-z0-9]?I\\b'
- '(?i)\\bN[^a-z0-9]?H[^a-z0-9]?L[^a-z0-9]?B[^a-z0-9]?I\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?A[^a-z0-9]?A[^a-z0-9]?A\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?A[^a-z0-9]?I[^a-z0-9]?D\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?A[^a-z0-9]?M[^a-z0-9]?S\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?B[^a-z0-9]?I[^a-z0-9]?B\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?C[^a-z0-9]?H[^a-z0-9]?D\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?D[^a-z0-9]?A\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?D[^a-z0-9]?C[^a-z0-9]?D\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?D[^a-z0-9]?C[^a-z0-9]?R\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?D[^a-z0-9]?D[^a-z0-9]?K\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?E[^a-z0-9]?H[^a-z0-9]?S\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?G[^a-z0-9]?M[^a-z0-9]?S\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?M[^a-z0-9]?H\\b'
- '(?i)\\bN[^a-z0-9]?IM[^a-z0-9]?H[^a-z0-9]?D\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?N[^a-z0-9]?D[^a-z0-9]?S\\b'
- '(?i)\\bN[^a-z0-9]?I[^a-z0-9]?N[^a-z0-9]?R\\b'
- '(?i)\\bN[^a-z0-9]?E[^a-z0-9]?I\\b' AND 'EY[^a-z0-9]?[0-9]{4,7}' OR (('(?i)\\bTW[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bTR[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bAT[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bCA[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bEY[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bHG[^a-z0-9]?[0-9]{4,7}' OR 'HL[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bAG[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bAA[^a-z0-9]?[0-9]{4,7}' OR 'AI[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bAR[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bEB[^a-z0-9]?[0-9]{4,7}' OR 'HD[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bDA[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bDC[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bDE[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bDK[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bES[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bGM[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bMH[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bMD[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bNS[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bNR[^a-z0-9]?[0-9]{4,7}' OR '(?i)\\bLM[^a-z0-9]?[0-9]{4,7}') AND ('(?i)\\bRO1' OR '(?i)\\bRO3' OR '(?i)\\bR13' OR '(?i)\\bR15' OR '(?i)\\bR21' OR '(?i)\\bR34' OR '(?i)\\bR41' OR '(?i)\\bR42' OR '(?i)\\bR44' OR '(?i)\\bR56' OR '(?i)\\bU01' OR '(?i)\\bK99' OR '(?i)\\bPO1' OR '(?i)\\bP20' OR '(?i)\\bP30' OR '(?i)\\bP50' OR '(?i)\\bR24' OR '(?i)\\bR25' OR '(?i)\\bXO1' OR '(?i)\\bNational inst' OR '(?i)\\bNat[^a-z0-9]{0,2}inst' OR '(?i)\\bNatl[^a-z0-9]{0,2}inst' OR '(?i)\\bNIH'))

National Research Foundation Singapore

- '(?i)National[^a-z0-9]?Research[^a-z0-9]?Found' AND '(?i)Singapore'
- '(?i)\\bNat[^a-z0-9]{0,2}Res[^a-z0-9]{0,2}Found' AND '(?i)Singapore'
- '(?i)\\bN[^a-z0-9]{0,1}R[^a-z0-9]{0,1}F\\b' AND '(?i)Singapore'

National Science Centre, Poland

- '(?i)National[^a-z0-9]?Science[^a-z0-9]?Cent' AND ('(?i)Poland' OR '(?i)Polish')

National Science and Engineering Research Council of Canada

- '(?i)Natural[^a-z0-9]?Science[a-z]?[^a-z0-9]?(and|&)[^a-z0-9]?Engineering[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Conseil[^a-z0-9]?de[^a-z0-9]?recherche[a-z]?[^a-z0-9]?en[^a-z0-9]?science[a-z]?[^a-z0-9]?naturelle[a-z]?[^a-z0-9]?et[^a-z0-9]?en[^a-z0-9]?g.nie'
- '(?i)\\bN[^a-z0-9]{0,1}S[^a-z0-9]{0,1}E[^a-z0-9]{0,1}R[^a-z0-9]{0,1}C\\b'
- '(?i)\\bC[^a-z0-9]{0,1}R[^a-z0-9]{0,1}S[^a-z0-9]{0,1}N[^a-z0-9]{0,1}G\\b'

Project HOPE

- '(?i)\\bProject[^a-z0-9]?H[^a-z0-9]{0,1}O[^a-z0-9]{0,1}P[^a-z0-9]{0,1}E\\b'

Research Council, Norway

- '(?i)R[^a-z0-9]{0,1}C[^a-z0-9]{0,1}N[^a-z0-9]{0,1}N[^a-z0-9]{0,1}F[^a-z0-9]{0,1}R\\b'
- ('(?i)Research[^a-z0-9]?Council' OR '(?i)\\bN[^a-z0-9]{0,1}F[^a-z0-9]{0,1}R\\b') AND ('(?i)Norway' OR '(?i)Norwegian')

Science Europe

- '(?i)\\bScience[^a-z0-9]?Europe\\b'

Science Foundation, Ireland

- '(?i)Science[^a-z0-9]?Found' AND '(?i)Ireland'
- '\\bS[^a-z0-9]{0,1}F[^a-z0-9]{0,1}I\\b' AND '(?i)Ireland'
- '(?i)Science[^a-z0-9]?Found' AND '(?i)Irish'
- '\\bS[^a-z0-9]{0,1}F[^a-z0-9]{0,1}I\\b' AND '(?i)Irish'

South African Medical Research Council

- ('(?i)Medical[^a-z0-9]?Research[^a-z0-9]?Council' AND '(?i)South Africa')
- '(?i)\\bS[^a-z0-9]{0,1}A[^a-z0-9]{0,1}M[^a-z0-9]{0,1}R[^a-z0-9]{0,1}C\\b'
- '(?i)bAfrican M[^a-z0-9]{0,1}R[^a-z0-9]{0,1}C\\b'
- '(?i)bAfrica M[^a-z0-9]{0,2}R[^a-z0-9]{0,2}C\\b'

Swedish Research Council

- '(?i)Research[^a-z0-9]?Council' AND ('(?i)Swedish' OR '(?i)Sweden')

UK Department for International Development

- '(?i)U[^a-z0-9]{0,2}K[^a-z0-9]{0,2}Department[^a-z0-9]?for[^a-z0-9]?International[^a-z0-9]?Dev'
- '(?i)\\bD[^a-z0-9]{0,2}F[^a-z0-9]{0,2}I[^a-z0-9]{0,2}D\\b'

Institut Pasteur

- '(?i)Institut[^a-z0-9]?Pasteur'
- '(?i)Pasteur[^a-z0-9]?Institut'
- '(?i)Pasteur[^a-z0-9]?Covid.?19'

The Royal Society

- '(?i)Royal[^a-z0-9]?Society'

UK Research and Innovation

- '(?i)\\bU[^a-z0-9]{0,1}K[^a-z0-9]{0,1}Research and Innovation'

- '(?i)\\bResearch and Innovation[^a-z0-9]{0,1}U[^a-z0-9]{0,1}K'
- '(?i)\\bResearch[^a-z0-9]?(and|&)[^a-z0-9]?Innovation[^a-z0-9]{0,1}U[^a-z0-9]{0,1}K'
- '(?i)Research[^a-z0-9]?Council[a-z]?[^a-z0-9]?U[^a-z0-9]{0,1}K\\b'
- '(?i)\\bU[^a-z0-9]{0,1}K[^a-z0-9]{0,1}Research[^a-z0-9]?Council'
- '(?i)Arts[^a-z0-9]?(and|&)[^a-z0-9]?Humanities[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Arts[^a-z0-9]?(and|&)[^a-z0-9]?Humanities R[^a-z0-9]{0,1}C'
- '(?i)Biotechnology[^a-z0-9]?(and|&)[^a-z0-9]?Biological[^a-z0-9]?Sciences[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Biotechnology[^a-z0-9]?(and|&)[^a-z0-9]?Biological[^a-z0-9]?Sciences[^a-z0-9]?R[^a-z0-9]{0,1}C'
- '(?i)\\bB[^a-z0-9]?B[^a-z0-9]?S[^a-z0-9]?R[^a-z0-9]?C\\b'
- '(?i)Economic[^a-z0-9]?(and|&)[^a-z0-9]?Social[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Economic[^a-z0-9]?(and|&)[^a-z0-9]?Social R[^a-z0-9]{0,2}C'
- '(?i)\\bE[^a-z0-9]?S[^a-z0-9]?R[^a-z0-9]?C\\b'
- '(?i)Engineering[^a-z0-9]?(and|&)[^a-z0-9]?Physical[^a-z0-9]?Sciences[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Engineering[^a-z0-9]?(and|&)[^a-z0-9]?Physical[^a-z0-9]?Sciences[^a-z0-9]?R[^a-z0-9]{0,1}C'
- '(?i)\\bE[^a-z0-9]?P[^a-z0-9]?S[^a-z0-9]?R[^a-z0-9]?C'
- '(?i)Natural[^a-z0-9]?Environment[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Natural[^a-z0-9]?Environment[^a-z0-9]?R[^a-z0-9]{0,1}C'
- '(?i)\\bN[^a-z0-9]?E[^a-z0-9]?R[^a-z0-9]?C\\b'
- (?i)Natural[^a-z0-9]?Environment[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Natural[^a-z0-9]?Environment[^a-z0-9]?R[^a-z0-9]{0,1}C'
- '(?i)\\bN[^a-z0-9]?E[^a-z0-9]?R[^a-z0-9]?C\\b'
- '(?i)\\bCentre[^a-z0-9]?for[^a-z0-9]?the[^a-z0-9]?Replacement[^a-z0-9]{0,2}Refinement\\b'
- '(?i)\\bN[^a-z0-9]?C[^a-z0-9]?3[^a-z0-9]?R[^a-z0-9]?s\\b'
- '(?i)Innovate[^a-z0-9]?U[^a-z0-9]{0,1}K'
- '(?i)Research England'
- '(?i)Science[^a-z0-9]?(and|&)[^a-z0-9]?Technology[^a-z0-9]?Facilities[^a-z0-9]?Council'
- '(?i)\\bS[^a-z0-9]?T[^a-z0-9]?F[^a-z0-9]?C\\b'
- '(?i)\\bU[^a-z0-9]?K[^a-z0-9]?R[^a-z0-9]?I\\b'
- (?i)U[^a-z0-9]{0,1}K[^a-z0-9]{0,2}Medical[^a-z0-9]?Research[^a-z0-9]?Council'
- '(?i)Medical[^a-z0-9]?Research[^a-z0-9]?Council[^a-z0-9]{0,2}U[^a-z0-9]{0,1}K'
- '(?i)Medical[^a-z0-9]?Research[^a-z0-9]?Council[^a-z0-9]{0,2}M[^a-z0-9]?R[^a-z0-9]?C\\b'
- '(?i)\\bU[^a-z0-9]?K[^a-z0-9]{0,2}M[^a-z0-9]?R[^a-z0-9]?C'
- '(?i)\\bM[^a-z0-9]?R[^a-z0-9]?C[^a-z0-9]{0,2}U[^a-z0-9]?K'

Volkswagen Foundation

- '(?i)Volkswagen[^a-z0-9]?Found'
- '(?i)VolkswagenStiftung'

Wellcome Trust

- '(?i)Wellcome'

XPRIZE Fondation

- '(?i)XPRIZE[^a-z0-9]?fond'

Netherlands Organisation for Health Research and Development

- '(?i)Organisation for Health Research and Development|Organization for Health Research and Development' AND '(?i)Netherlands'
- '(?i)organisatie[^a-z0-9]?voor[^a-z0-9]?gezondheidsonder\\B'
- '(?i)Zon[^a-z0-9]{0,1}M[^a-z0-9]{0,2}W'

The following search terms were employed in funding acknowledgement sections to identify support by Zika Joint Statement signatory funders (that were not signatories to the Covid-19 Statement as well):

Chinese Academy of Sciences

- ('(?i)\\bC[^a-z0-9]{0,1}A[^a-z0-9]{0,1}S\\b' AND ('(?i)china' OR '(?i)chinese')
- '(?i)\\bChinese[^a-z0-9]?Academy[^a-z0-9]?of[^a-z0-9]?Sci'
- '(?i)\\bChinese[^a-z0-9]?Acad[^a-z0-9]{0,2}of Sci'
- '(?i)\\bChinese[^a-z0-9]?Acad[^a-z0-9]{0,2}Sci'
- '(?i)\\bChina[^a-z0-9]?Academy[^a-z0-9]?of[^a-z0-9]?Sci'
- '(?i)\\bChina[^a-z0-9]?Acad[^a-z0-9]{0,2}of Sci'
- '(?i)\\bChina[^a-z0-9]?Acad[^a-z0-9]{0,2}Sci'

Deutsche Forschungsgemeinschaft

- '(?i)\\bD[^a-z0-9]{0,1}F[^a-z0-9]{0,1}G\\b'
- '(?i)\\bDeutsche[^a-z0-9]?Forschungsgemeins'
- '(?i)\\bGerman[^a-z0-9]?research found'
- '(?i)\\bGerman[^a-z0-9]?res[^a-z0-9]{0,2}found'

Oswaldo Cruz Foundation

- '(?i)Oswaldo[^a-z0-9]?Cruz[^a-z0-9]?Found'
- '(?i)funda..o Oswaldo[^a-z0-9]?Cruz'
- '(?i)foundation[^a-z0-9]?Oswaldo[^a-z0-9]?Cruz'
- '(?i)\\bFiocruz\\b'

US National Science Foundation

- (('(?i)\\bN[^a-z0-9]?S[^a-z0-9]?F\\b' OR '(?i)\\bNational Science Foundation\\b' OR '(?i)\\bN[^a-z0-9]?S[^a-z0-9]?F(\\W+\\w+){0,4} U[^a-z0-9]?S' OR '(?i)\\bU[^a-z0-9]?S(\\W+\\w+){0,4} N[^a-z0-9]?S[^a-z0-9]?F' OR '(?i)\\bN[^a-z0-9]?S[^a-z0-9]?F(\\W+\\w+){0,4} United[^a-z0-9]?states' OR '(?i)\\bUnites States(\\W+\\w+){0,4} N[^a-z0-9]?S[^a-z0-9]?F')

AND NOT

('(?i)\\bSwiss N[^a-z0-9]?S[^a-z0-9]?F' OR '(?i)\\bSwiss National Science Foundation' OR '(?i)\\bN[^a-z0-9]?S[^a-z0-9]?F[^a-z0-9]?C\\b' OR '(?i)\\bN[^a-z0-9]?N[^a-z0-9]?S[^a-z0-9]?F[^a-z0-9]?C\\b' OR '(?i)\\bN[^a-z0-9]?S[^a-z0-9]?F(\\W+\\w+){0,7} china' OR '(?i)\\bN[^a-z0-9]?S[^a-z0-9]?F(\\W+\\w+){0,7} chinese' OR '(?i)\\bChina(\\W+\\w+){0,7} N[^a-z0-9]?S[^a-z0-9]?F\\b' OR '(?i)\\bChinese(\\W+\\w+){0,7} N[^a-z0-9]?S[^a-z0-9]?F\\b' OR '(?i)\\bScience Foundation(\\W+\\w+){0,7} china' OR '(?i)\\bScience Foundation(\\W+\\w+){0,7} chinese' OR '(?i)\\bChina(\\W+\\w+){0,7} National Science Found' OR '(?i)\\bChinese(\\W+\\w+){0,7} National Science Found' OR '(?i)\\bN[^a-z0-9]?N[^a-z0-9]?S[^a-z0-9]?F(\\W+\\w+){0,7} china' OR '(?i)\\bN[^a-z0-9]?N[^a-z0-9]?S[^a-z0-9]?F(\\W+\\w+){0,7} chinese' OR '(?i)\\bChina(\\W+\\w+){0,7} N[^a-z0-9]?N[^a-z0-9]?S[^a-z0-9]?F\\b' OR (?i)\\bChinese(\\W+\\w+){0,7} N[^a-z0-9]?N[^a-z0-9]?S[^a-z0-9]?F\\b' OR '(?i)\\bProvince(\\W+\\w+){0,7} N[^a-z0-9]?S[^a-z0-9]?F\\b' OR '(?i)\\bN[^a-z0-9]?S[^a-z0-

9]?F(\\W+\\w+){0,7} province\\b' OR '(?i)\\bNational(\\W+\\w+){0,6} N[^a-z0-9]?S[^a-z0-9]?F\\b'))

The following funders were also signatories of the COVID-19 Statement and therefore their queries are not repeated here:

- Bill & Melinda Gates Foundation
- Canadian Institutes of Health Research
- Chinese Centre for Disease Control and Prevention
- Department of Biotechnology, India
- Doctors Without Borders
- Fondation Mérieux
- Institut Pasteur
- Japan Agency for Medical Research and Development
- Netherlands Organisation for Health Research and Development
- South African Medical Research Council
- UK Academy of Medical Sciences
- UK Biotechnology and Biological Sciences Research Council
- UK Department for International Development
- UK Medical Research Council
- US Centers for Disease Control and Prevention
- US National Academy of Medicine
- US National Institutes of Health
- Wellcome Trust

Appendix F: Regex queries for assembling the DAS excerpt set

The following search terms were employed in the regex queries to identify and extract text excerpts from the DAS sections.

- '(?i)(?s)((.){0,20}[^a-z]([^a-z]Data[^a-z]{0,1}Sharing[^a-z]{0,1}(?!platform))[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z]([^a-z]Data[^a-z]{0,1}Availability)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}Deposition)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}Archiving)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}Accessibility)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z]([^a-z]Availability[^a-z]{0,1}of[^a-z]{0,1}Data)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Deposition[^a-z]{0,1}of[^a-z]{0,1}Data)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Archiving[^a-z]{0,1}of[^a-z]{0,1}Data)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Accessibility[^a-z]{0,1}of[^a-z]{0,1}Data)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}Sharing)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}Availability)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}Deposition)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}Archiving)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}Accessibility)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Availability[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}code)[^a-z](.){0,800})'

- '(?i)(?s)((.){0,20}[^a-z](Deposition[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}code)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Archiving[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}code)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Accessibility[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}code)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}Sharing)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?[^a-z]{0,1}Availability)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?[^a-z]{0,1}Deposition)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?[^a-z]{0,1}Archiving)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?[^a-z]{0,1}Accessibility)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Availability[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Deposition[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Archiving[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Accessibility[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}material[s]?)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}software[^a-z]{0,1}Availability)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}software[^a-z]{0,1}Deposition)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}software[^a-z]{0,1}Archiving)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}and[^a-z]{0,1}software[^a-z]{0,1}Accessibility)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Availability[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}software)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Deposition[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}software)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Archiving[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}software)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Accessibility[^a-z]{0,1}of[^a-z]{0,1}Data[^a-z]{0,1}and[^a-z]{0,1}software)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Deposited[^a-z]{0,1}data)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z]((?!must include a data)Availability[^a-z]{0,1}Statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Deposition[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](archiving[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](accessibility[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](open[^a-z]{0,1}data[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](open[^a-z]{0,1}data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](open[^a-z]{0,1}data[^a-z]{0,1}and[^a-z]{0,1}materials[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](Data[^a-z]{0,1}statement)[^a-z](.){0,800})'

- '(?i)(?s)((.){0,20}[^a-z](data[^a-z]{0,1}and[^a-z]{0,1}code[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](data[^a-z]{0,1}and[^a-z]{0,1}material[s]?[^a-z]{0,1}statement)[^a-z](.){0,800})'
- '(?i)(?s)((.){0,20}[^a-z](data[^a-z]{0,1}and[^a-z]{0,1}software[^a-z]{0,1}statement)[^a-z](.){0,800})'

Appendix G: Regex queries for identifying data deposition mentions within the DAS mentions

The following search terms were employed within the DAS excerpts to identify deposits of datasets to open repositories from the selected list.

- '(?s)(.){0,200}[^a-z](MassIVE)[^a-z](.){0,200}'
- '(?s)(.){0,200}[^a-z](IntAct)[^a-z](.){0,200}'
- '(?s)(.){0,200}[^a-z](PRIDE)[^a-z](.){0,200}'
- '(?s)(.){0,200}[^a-z](CAPACITY)[^a-z](.){0,200}'
- '(?s)(.){0,200}[^a-z](ENERGY)[^a-z](.){0,200}'
- '(?s)(.){0,200}[^a-z](PRIORITY)[^a-z](.){0,200}' ‘
- '(?i)(?s)(.){0,200}[^a-z](cif)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cardiac complications in patients with sars)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NDAR)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ENA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GEO)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NDEx)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ITIS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](github)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Genbank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PDB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SRA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](accession number)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](R code)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](accession numbers)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Protein Data Bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](source code)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](osf)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Gene Expression Omnibus)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](python script)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](BioProject)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](R script)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](accession code)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](zenodo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](National Center for Biotechnology Information)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Sequence Read Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](figshare)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](accession codes)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](open science framework)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](CCDC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UniProtKB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Synapse)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](ProteomeXchange)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Electron microscopy data[^a-z]{0,1}bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](emdb)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](python code)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](European Nucleotide Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PubChem)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EMBL-EBI)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dataverse)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](analysis script)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](matlab code)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ArrayExpress)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dryad)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](MIT license)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GPL license)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](matlab script)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dbSNP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dbGaP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](VectorBase)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Protein Databank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ChEMBL)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](mendeley data)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Array Express)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UK Data Service)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Influenza Research Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ClinVar)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](European Molecular Biology Laboratory)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ImmPort)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](wwPDB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](FlowRepository)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GBIF)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](DDBJ)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](BMRB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NDCT)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Global Biodiversity Information Facility)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](metabolights)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](MetaboLights)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](DNA Data Bank of Japan)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NCBI Assembly)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Mouse Genome Informatics)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](The Network Data Exchange)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Cambridge Crystallographic Data Centre)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PhysioNet)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Worldwide Protein Data Bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](MG-RAST)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Proteome Exchange)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Cancer Imaging Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PeptideAtlas)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EBRAINS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Biological Magnetic Resonance Data Bank)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](G-Node)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Mgnify)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Image Data Resource)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Treebase)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ZFIN)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](HIV Data Archive Program)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NAHDAP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ProteomeExchange)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Biological General Repository for Interaction Datasets)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GIGADB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NITRC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Integrated Taxonomic Information System)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](European Variation Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](OpenNeuro)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](STRENDA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GigaScience)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](10.1759)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](10.5073)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](10.25493)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](10.6073)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](10.15468)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](10.5063)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](iu.bl)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](edi.edi)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](crossref.citations)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](pds.data)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dk.ku)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](usc.isi)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](usc.ini)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](lg.lg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.av)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](inist.health)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.eap)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.ed)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dot.dot)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PANGAEA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](und.library)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ada.repo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](tib.si)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](APOLLO)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.mmu)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.dri)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](nrc.ir)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](aad.repo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dk.sa)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](fct.unl)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cern.yellow)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](fct.st)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](iu.sd)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](carr.carr)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](idiap.idiap)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](tib.eta)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.hi)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](nist.admin)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ala.repo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.at)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.repo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](si.si)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](gt.ir)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](nasa.data)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rit.rit)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.open)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.labs)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cul.columbia)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dk.sb)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.lse)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](osf.osf)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](inist.ca)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](sab.sab)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.ar)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ieee.dataport)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](au.dra)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](anu.repo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](si.cda)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](snd.snd)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.sh)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](USGS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rutgers.lib)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bl.gold)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.im)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](DATASPACE)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rice.kinder)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rads.de)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](gesis.wi)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](figshare.up)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](lbs.lbs)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](occ.occ)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](zbw.zew)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](AADC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ADS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](AEKOS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](AEKOS - TERN Ecoinformatics)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](AODN)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Archaeology Data Service)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Australian Antarctic Data Centre)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Australian Ocean Data Network)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](BGS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](BioGRID)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](Biological General Repository for Interaction Datasets)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Biological Magnetic Resonance Data Bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](BioModels Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](BMRB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](British Geological Survey)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](caNanoLab)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Cell Image Library)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](CIL)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ClinicalTrials.gov)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](COD)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Coherent X-ray Imaging Data[^a-z]{0,1}Bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Crystallography Open Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](CXIDB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Database of Interacting Proteins)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](datadryad.org)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dbGAP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dbSNP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dbVar)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](DIP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](DNA DataBank of Japan)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EarthChem)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EBRAINS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EDI)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EGA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EMDataBank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Environmental Data Initiative)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Eukaryotic Pathogen Database Resources)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EuPathDB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](European Genome-phenome Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](European Nucleotide Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](European Variation Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](EVA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](FlowRepository)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](FlyBase)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GBIF)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GEA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GenomeRNAi)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Genomic Expression Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GES DISC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GIN)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](G-Node)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](HEPData)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](IDR)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Image Data Resource)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Incorporated Research Institutions for Seismology (IRIS))[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Influenza Research Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Integrated Taxonomic Information System)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](IRIS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ITIS)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](Japanese Genotype-phenotype Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](JGA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](KiMoSys)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Kinetic Models of Biological Systems)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](KNB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](KNB: The Knowledge Network for Biocomplexity)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Magnetics Information Consortium (MagIC))[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Marine Data Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Marine Geosciences Data System)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Materials Cloud Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](MGDS)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](MGI)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](MorphoBank.org)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Mouse Genome Informatics)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Movebank Data Repository)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NAHDAP)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NASA Goddard Earth Sciences Data and Information Services Center)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](National Addiction & HIV Data Archive Program)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](National Database for Autism Research)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](National Database for Clinical Trials related to Mental Illness)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NCBI Assembly)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PubChem BioAssay)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PubChem Substance)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Sequence Read Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NCBI Taxonomy)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NCBI Trace Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NCEI)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NDAR)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NDCT)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NDEx)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NERC Data Centres)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Neuroimaging Informatics Tools and Resources Collaboratory (NITRC))[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NeuroMorpho.org)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NITRC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NOAA National Centers for Environmental Information)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NoMaD)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](NoMaD Repository)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Oak Ridge National Laboratory Distributed Active Archive Center)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Open Science Framework)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](openICSPR)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](OpenNeuro)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](OpenTopography)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ORNL DAAC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PCDDB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](PeptideAtlas)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](Protein Circular Dichroism Data Bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](QDR)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Qualitative Data Repository)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Rat Genome Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](RDoCdb)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Research Domain Criteria Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](RGD)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SBGrid)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SEANOE)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SICAS Medical Image Repository (formally Virtual Skeleton Database))[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SIMBAD Astronomical Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SMIR)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SRA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Structural Biology Data Grid)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](The Cancer Imaging Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](The Global Biodiversity Information Facility)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UK Data Service)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UK Solar System Data Centre)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UNAVCO)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UNAVCO, Inc.)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UniProtKB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](VectorBase)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](WDCC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](World Data Center for Climate at DKRZ)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Xenbase)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Zebrafish Model Organism Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ZFIN)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](fairsharing.org)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GISAID)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ENA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Japanese Genotype-phenotype Archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ScienceDB)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Science Data Bank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](federated research data)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](frdr)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](RSNA)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](RICORD)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Open Radiology data)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cohort collaborative)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](n3c)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](19 data platform)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](19 data portal)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](INSDC)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](international nucleotide sequence database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](RefSeq)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Reference Sequence Database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ncbi virus)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](gsa)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](genome sequence archive)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](UniProt)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Molecular interaction database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](GWAS Catalog)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](panorama public)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bndex)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](Computational chemistry dataset)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](bicpsr)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](skeleton database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](fairsharing)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cirr)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 imaging research)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](acs covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](college of surgeons covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](asco registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z]([^a-z]{0,1}19 in oncology registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](badbir)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dermatologists biology and)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ccc19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 and cancer consortium)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cibmtr)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 data collection)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covi-preg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 and pregnancy reg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 cvd registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 data sharing/br ini)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 global pediatric rheumatology)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid-hepatology)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coviddiab)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covims)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus and ms reporting dat)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](dermatology covid[^a-z]{0,1}19 reg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](universal study covid[^a-z]{0,1}19 reg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](discovery virus covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](elso)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](life support organization registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](neuro-covid registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](eular covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](rheumatology covid[^a-z]{0,1}19 registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](gcs-neurocovid)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](neurological dysfunction in covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 in pediatric cancer)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](hs-covid)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](suppurativa covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](leoss)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](lean european open survey)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 hcv co[^a-z]{0,1}infection)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](neatid)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](pediatric covid[^a-z]{0,1}19 case registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 hiv co[^a-z]{0,1}infection)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](psoprotect)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](psoriasis patient registry for outcomes)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](pregnancy coronavirus outcomes reg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SCMR covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](resonance covid[^a-z]{0,1}19 registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure[^a-z]{0,1}alopecia)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}alopecia)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure[^a-z]{0,1}celiac)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}celiac)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](SECURE-EoE/EGID)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}eosinophilic)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure-ibd)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}inflammatory)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure-liver)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure-cirrhosis)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}liver)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}cirrhosis)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure-scd)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](secure-va)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}sickle)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](coronavirus under research exclusion[^a-z]{0,1}vascular)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 in patients with td1)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](svin covid[^a-z]{0,1}19 registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](intervention neurology covid[^a-z]{0,1}19 registry)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](teravolt)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](thoracic cancers international covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 global rheumatology)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](uk paediatric oncology coronavirus)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](uk coronavirus cancer monitoring)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](ukccmp)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](echocardiography in covid[^a-z]{0,1}19)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](b1mg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](1+mg)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](1 million genomes)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](cngdb)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](china national genebank)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](sharing all influenza data)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 host genetics)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](hdruk)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](health data research uk innovation gateway)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](icoda)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 data alliance)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](iddo)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](infectious diseases data observatory)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](geneweaver)[^a-z](.){0,200}'

- '(?i)(?s)(.){0,200}[^a-z](database of genotypes and phenotypes)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](covid[^a-z]{0,1}19 and ms)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](qmenta imaging database)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](recodid)[^a-z](.){0,200}'
- '(?i)(?s)(.){0,200}[^a-z](reconciliation of cohort)[^a-z](.){0,200}'