# Who are the winners and loser of good data practices?

**OSC Data Event, 23rd Nov 2020**

Georgie Humphreys, Open Research, Wellcome

g.humphreys@wellcome.org

@GS_Humphreys

ORCID: 0000-0001-9947-0034

# Content

- Wellcome’s commitment
- Examples of the landscape of data sharing;
  - data availability statements
  - databases
- What more can funders do?
- Winners and losers?

# Wellcome's commitment to Open Research

- Wellcome is a global charitable foundation dedicated to supporting science to solve the urgent health challenges facing everyone.
- We want the outputs of the research we fund to be accessed and used in ways that maximise benefit to health & society
- Passionate champion and advocate of open access and data sharing, with long-standing policies in place (open access and data, software and materials management and sharing, increased expectations for rapid sharing where there is a public health imperative)

# Covid-19 data sharing

- Wellcome statement of 31 January 2020 - endorsed by over 150 organisations
- Called on **publishers** to ensure the research they published had clear data availability statements
- Called on **researchers** to share interim and final data as rapidly as possible

**Press release** | 31 January 2020

## Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak

The outbreak of the novel coronavirus (COVID-19) represents a significant and urgent threat to global health.

We call 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.

We affirm the commitment to the principles set out in the 2016 Statement on data sharing in public health emergencies, and will seek to ensure that the World Health Organization (WHO) has rapid access to emerging findings that could aid the global response.

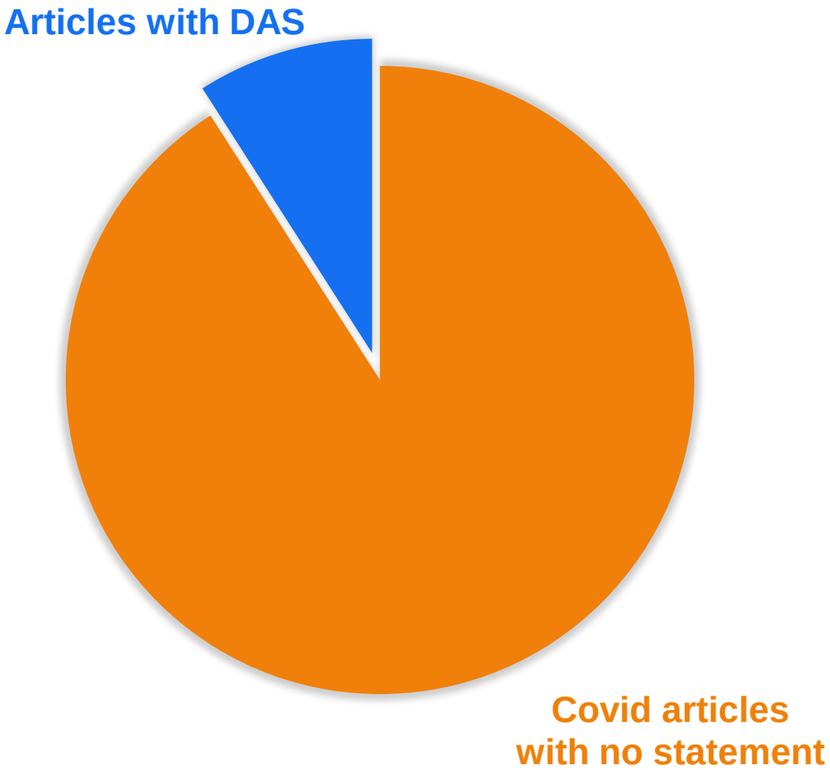
Specifically, we commit to work together to help ensure:

- all peer-reviewed research publications relevant to the outbreak are made immediately open access, or freely available at least for the duration of the outbreak
- research findings relevant to the outbreak are shared immediately with the WHO upon journal submission, by the journal and with author knowledge
- research findings are made available via preprint servers before journal publication, or via platforms that make papers openly accessible before peer review, with clear statements regarding the availability of underlying data
- researchers share interim and final research data relating to the outbreak, together with protocols and standards used to collect the data, as rapidly and widely as possible - including with public health and research communities and the WHO
- authors are clear that data or preprints shared ahead of submission will not pre-empt its publication in these journals

https://wellcome.org/coronavirus-covid-19/open-data

# Analysis of Data Availability Statements in Europe PMC

- Europe PMC has 66,294 full-text research articles, published in 2020, related to COVID-19
- Of these, just 6,180 (9%) have data availability statements
- Interestingly, of all full text articles in Europe PMC in 2020 (not just COVID articles), **23%** have Data Availability Statements
- 51 *Wellcome Open Research* articles on COVID-19 and 90% (46) have data availability statements




*Figures checked on 17/11/20*

# Lancet updates data availability requirements

## The Lancet changes editorial policy after hydroxychloroquine Covid study retraction

**New policy comes after serious quality control questions were raised about the data relied on by a study in the medical journal**

▲ The Lancet has changed its editorial policy after publishing a study in May which concluded that Covid-19 patients who received the drug hydroxychloroquine were dying at higher rates. But figures on the number of deaths and patients in hospital cited by the authors did not match up with official government and health department data. Photograph: George Frey/Reuters

One of the world's leading medical journals, the Lancet, has reformed its editorial policies following a shocking case of apparent research misconduct involving the study of hydroxychloroquine as a treatment for Covid-19.

COMMENT | ONLINE FIRST

## Learning from a retraction

The Editors of the Lancet Group

Published: September 17, 2020 • DOI: https://doi.org/10.1016/S0140-6736(20)31958-9 • Check for updates

PlumX Metrics

The publication and subsequent retraction[1, 2] in June, 2020, of the Article Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis, based on an alleged dataset associated with Surgisphere, prompted us to examine *The Lancet*'s peer-review processes to identify ways of further reducing risks of research and publication misconduct. As a result of this review, with immediate effect, we have made changes to the declarations we seek from authors, the data sharing statements we require for published research papers, and the peer-review process for similar papers based on large datasets or real-world data.

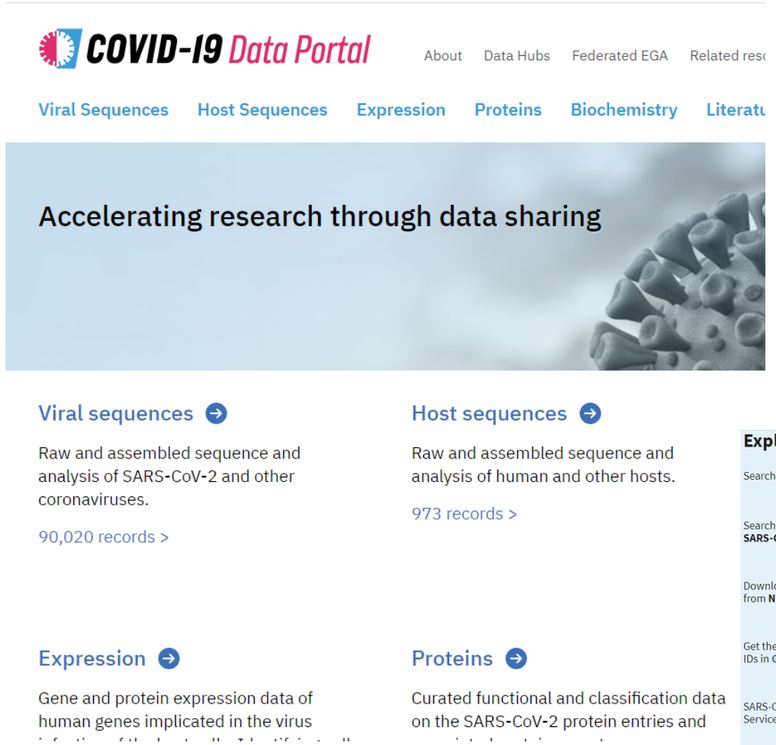
Changes to the signed declarations by authors in the author statements form will require that more than one author has directly accessed and verified the data reported in the manuscript. We will require that the authors who have accessed and verified underlying data are named in the contributors' statement. For research Articles that are the result of an academic and commercial partnership, one of the authors named as having accessed and verified data must be from the academic team. In addition, all authors will be asked to sign the author statements form to confirm they had full access to the data reported in their Article, and accept responsibility for submitting the Article for publication.

### Data sharing

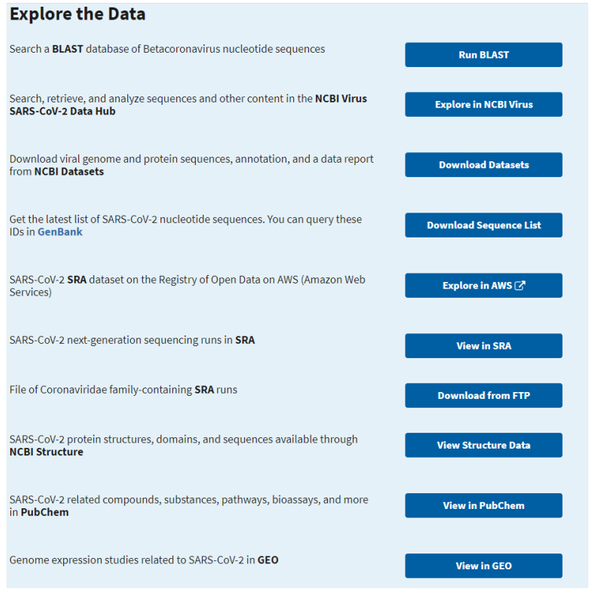
*The Lancet* journals are committed to open science, and require that all research papers include a data sharing statement from the authors that details what data will be shared, whether additional documents will be shared (e.g. the study protocol), when data will become available, and by what access criteria data will be shared. A data sharing statement will be published in the article.

Read a statement from the editors on data sharing statements for clinical trials (published June 2017) and a further update to our requirements on data access and sharing, published September 2020.

# …but COVID data are available for re-use




https://www.covid19dataportal.org/




https://www.ncbi.nlm.nih.gov/sars-cov-2/

## [Open Access] Download the Coronavirus Open Research Dataset

Download the COVID-19 Open Research Dataset, an extensive machine-readable full text resource of scientific literature with tens of thousands of articles about coronavirus.

Download CORD-19

https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge


https://search.vivli.org/?search=COVID%20or%20SARS-CoV-2%20or%202019-nCoV

# Level of sharing from research databases

Study of all research databases listed on the UK Health Research Authority's Assessment Review Portal that had received a favourable ethics opinion as of January 2018.

**Less than half (34%)** of research databases (n=354) **had granted access to their data and produced secondary publications.**

*"Time and again when questioned about this the excuse given was simply lack of resources/expertise/planning for database maintenance."*

*"The answer, I think, is for funders to commit to long term support for maintaining databases"*

*Simon Kolstoe (co-author)*

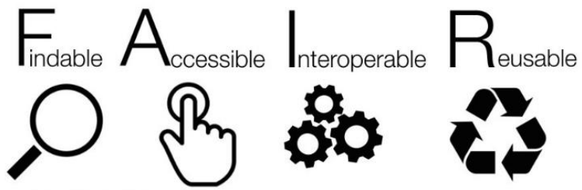


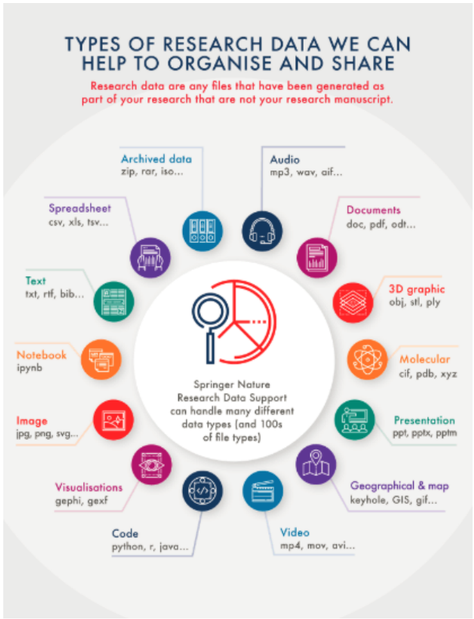
*BMJ Open 2020 Sep 30;10(9):e039756. doi: 10.1136/bmjopen-2020-039756*

# Funding and sustaining open research data

- implement our policy on data, software and materials management and sharing:
  - require outputs management plan (OMP) in every application
  - commit to support the costs associated with OMPs as an integral part of funding the research
- provide major funding for databases, tools and other key resources
- participate actively in key international initiatives to ensure sustained funding for core resources for life sciences research (e.g. ELIXIR and Global Biodata Coalition)
- Institutions to support researchers in meeting policy expectations and change incentives and rewards (including through implementing DORA)

# Supporting FAIR data




- Journals on supporting researchers in making data open and FAIR:
  - ongoing pilot with Springer-Nature to help researchers make data underlying papers findable and usable
  - our Wellcome Open Research publishing platform to link articles to underlying data and code

- With other funders - align policy expectations and develop automated approaches for assessing FAIR

# Winners and losers?

**g.humphreys@wellcome.org**