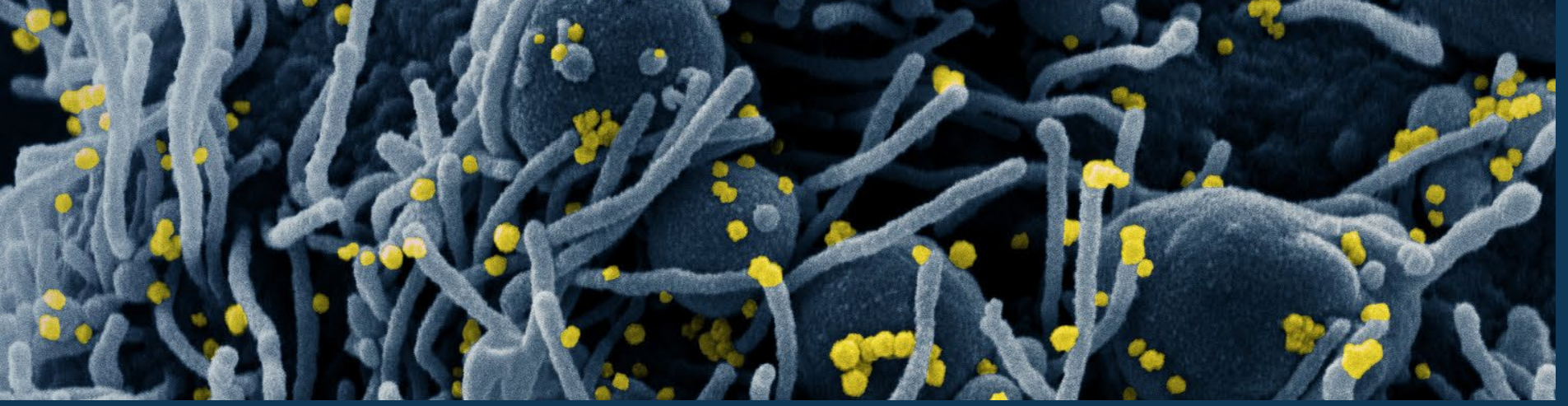
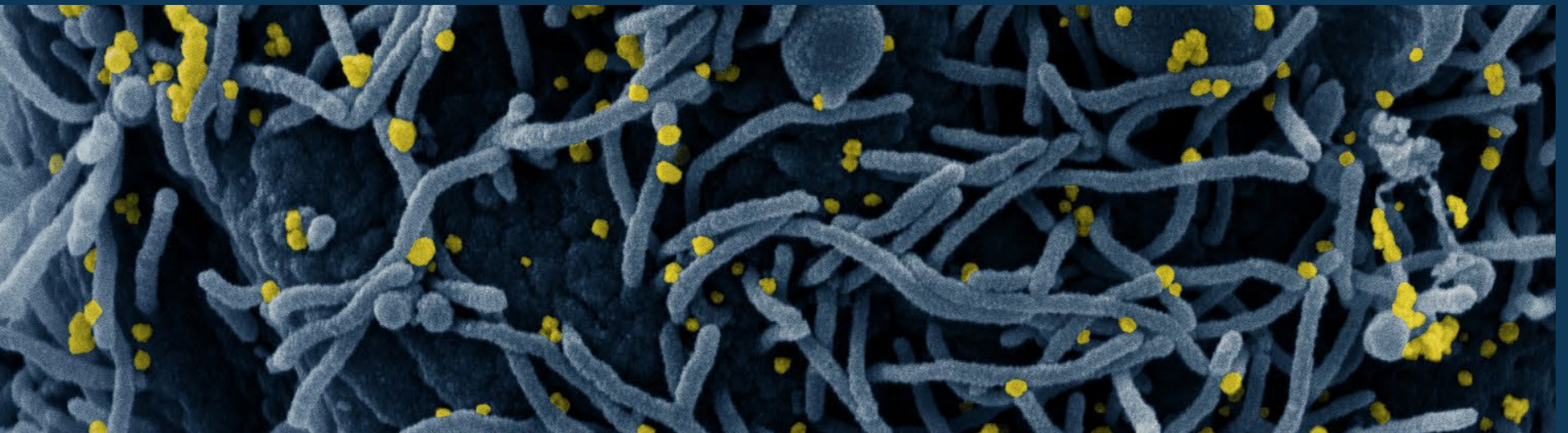


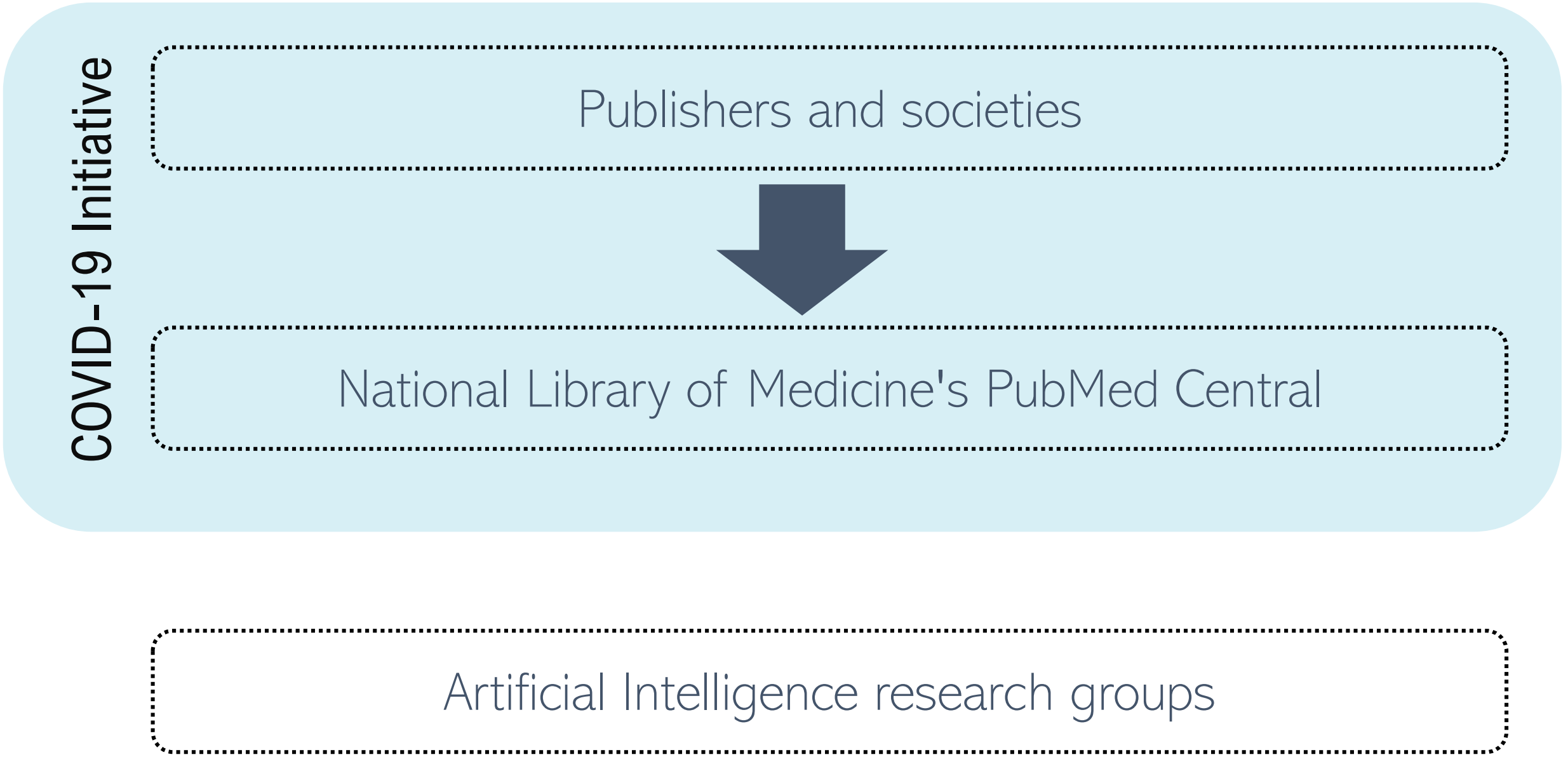
PubMed Central & COVID-19

Leveraging the literature as a dataset during a pandemic



Kathryn Funk, MLIS
Program Manager, PMC
April 24, 2020

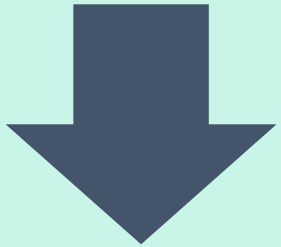
Credit: National Institute of Allergy and Infectious Diseases, NIH


COVID-19 Initiative
Publishers and societies
National Library of Medicine's PubMed Central
Artificial Intelligence research groups

Publishers and societies

CORD-19

National Library of Medicine's PubMed Central



Artificial Intelligence research groups

The countries listed below

urge publishers to voluntarily agree to make

their COVID-19 and coronavirus-related

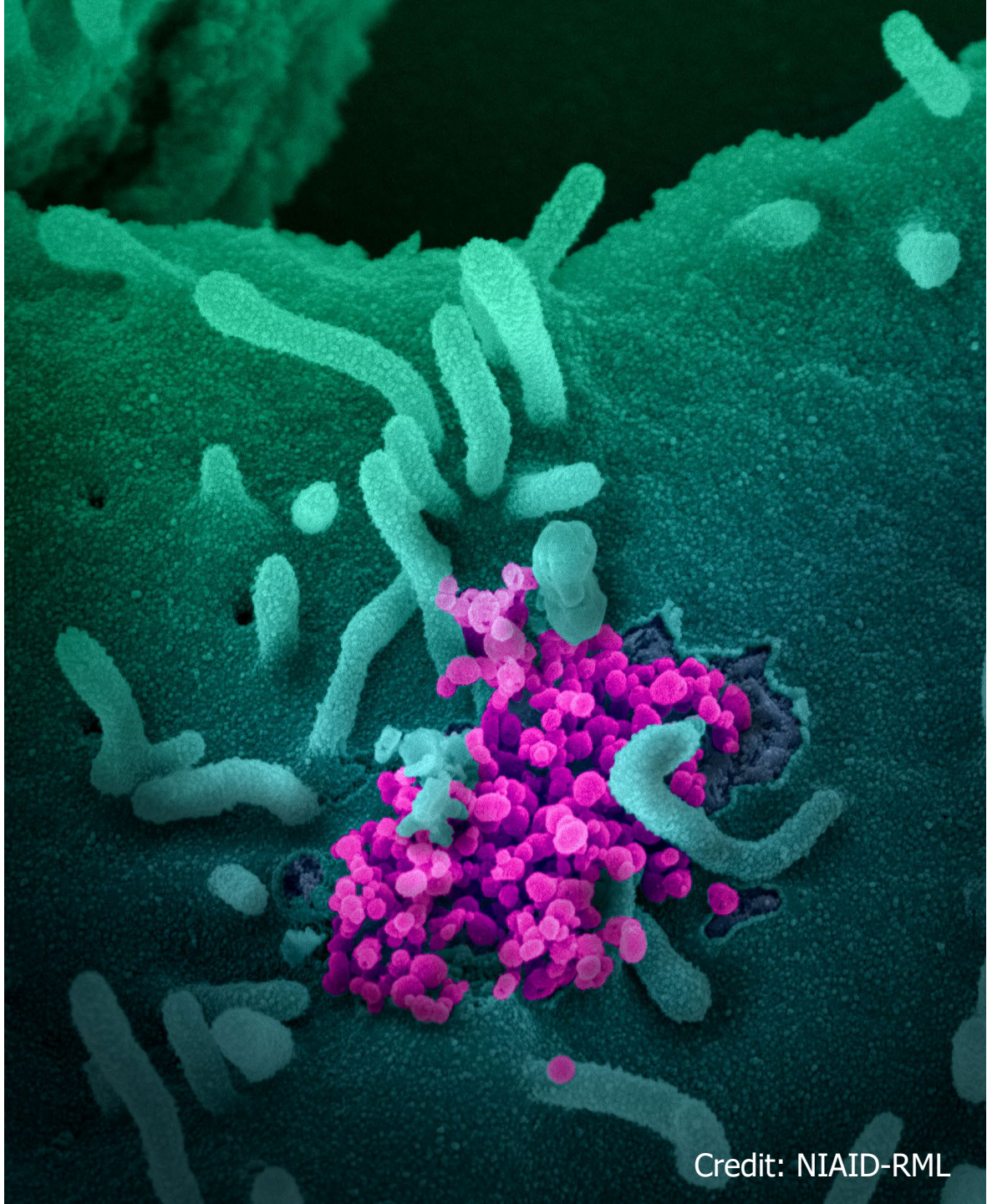
publications, and the available data supporting

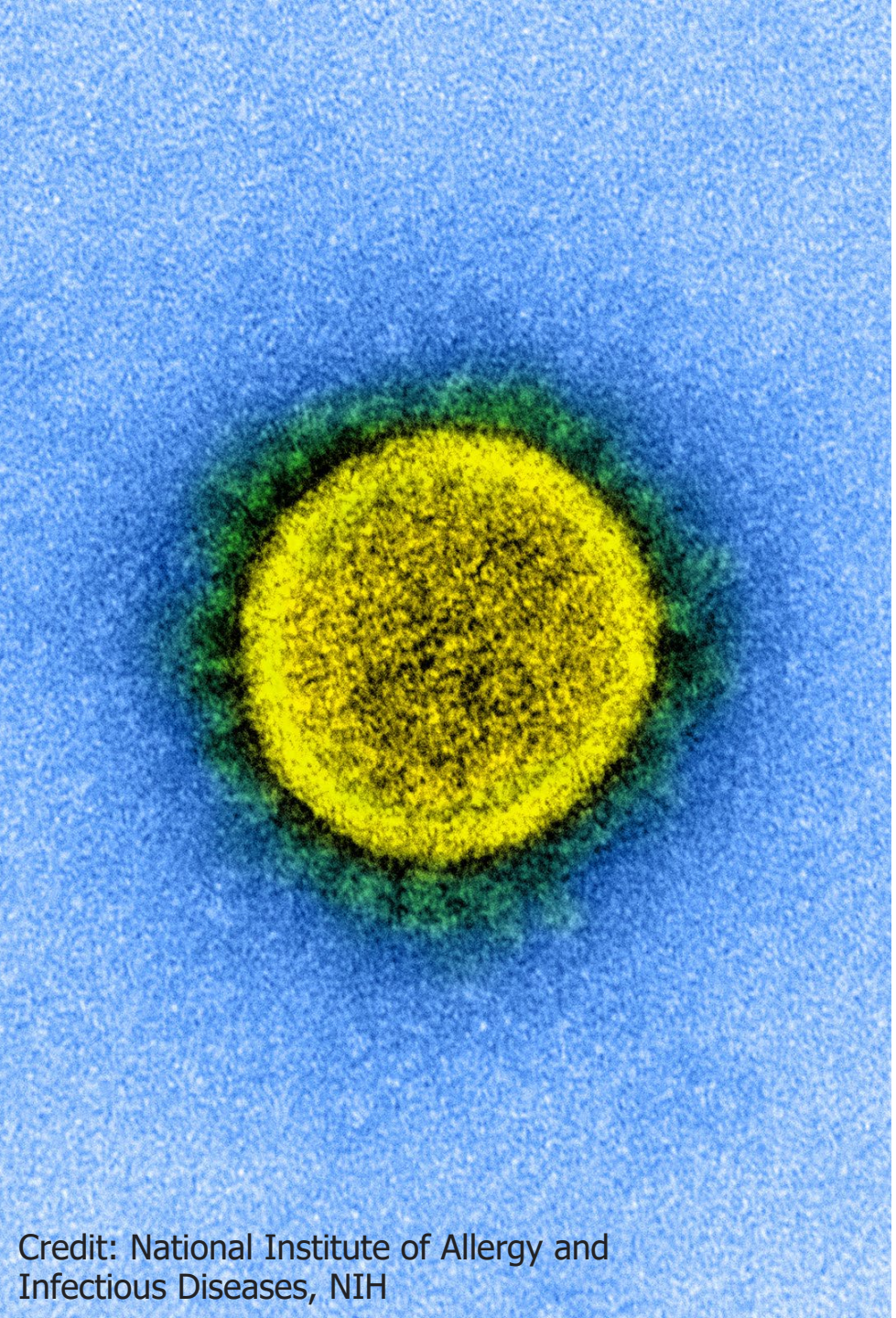
them, immediately accessible in PubMed

Central and other appropriate public repositories

... to support the ongoing public health

emergency response efforts.





Credit: National Institute of Allergy and Infectious Diseases, NIH

Scope of COVID-19 Initiative

Subject: Publications on COVID-19 and coronaviruses, more broadly.

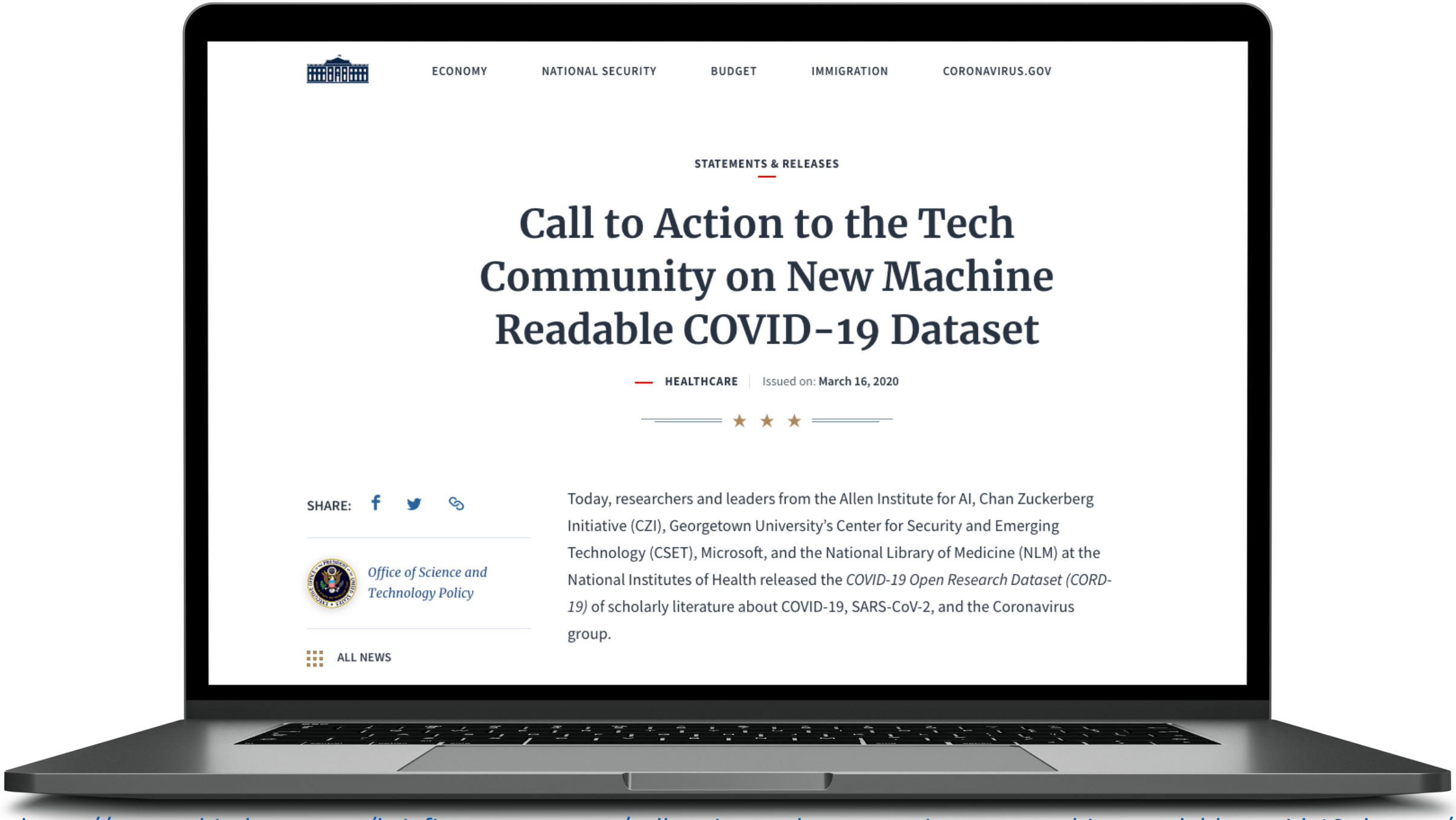
Date Range: Applies to articles published to date as well as future articles for the duration of this crisis.

Format: Human- and machine-readable formats

License: Must allow for research re-use and secondary analysis

Covid-19 Initiative To Date

- Nearly 50 publishers participating
- More than 38,000 coronavirus-related articles deposited with licenses that allow re-use and secondary analysis
- More than 3,000 COVID-19 specific articles deposited



https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19-dataset/

Covid-19 Open Research Dataset (CORD-19)

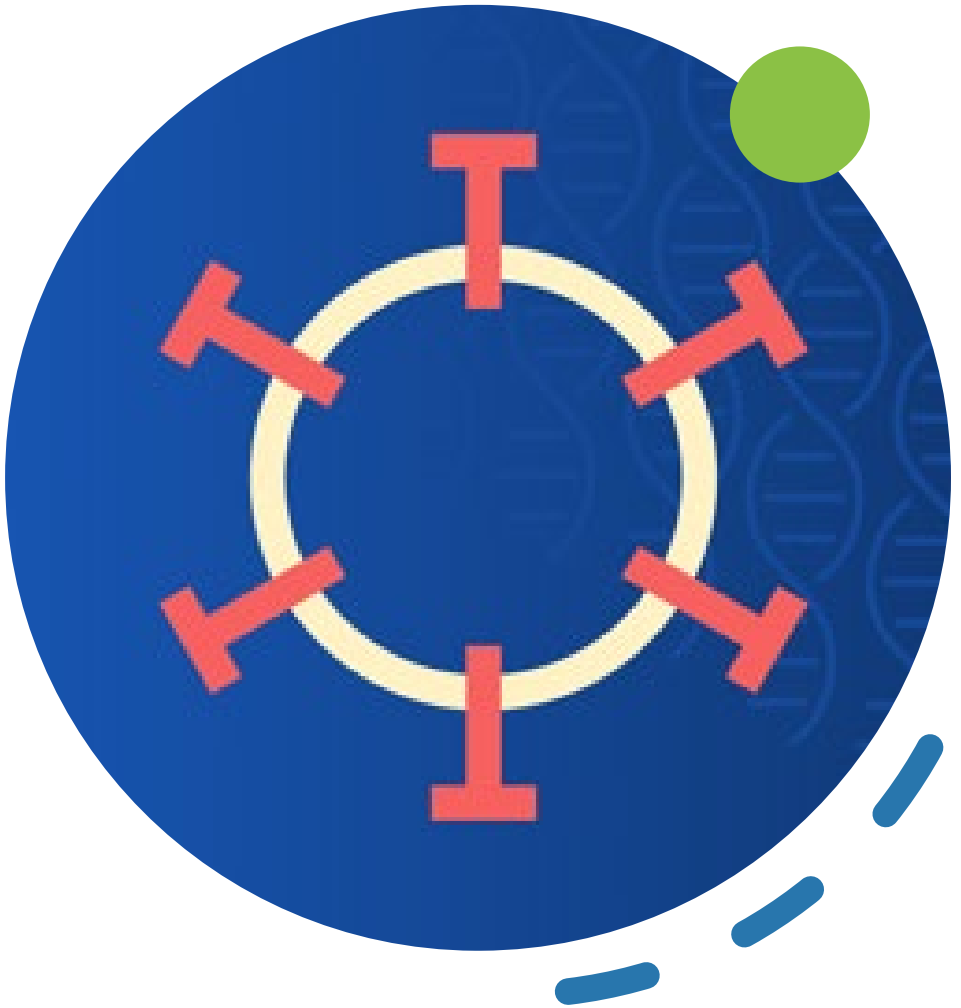
https://pages.semanticscholar.org/coronavirus-research

Scope of CORD-19

The dataset contains all COVID-19 and coronavirus-related research (e.g. SARS, MERS, etc.) from the following sources:

- NLM's PMC open access subset
- Additional COVID-19 research articles from a corpus maintained by the WHO
- bioRxiv and medRxiv pre-prints

Also provides a metadata file of 51,078 articles with links to PubMed, Microsoft Academic and the WHO COVID-19 database of publications.



https://pages.semanticscholar.org/coronavirus-research

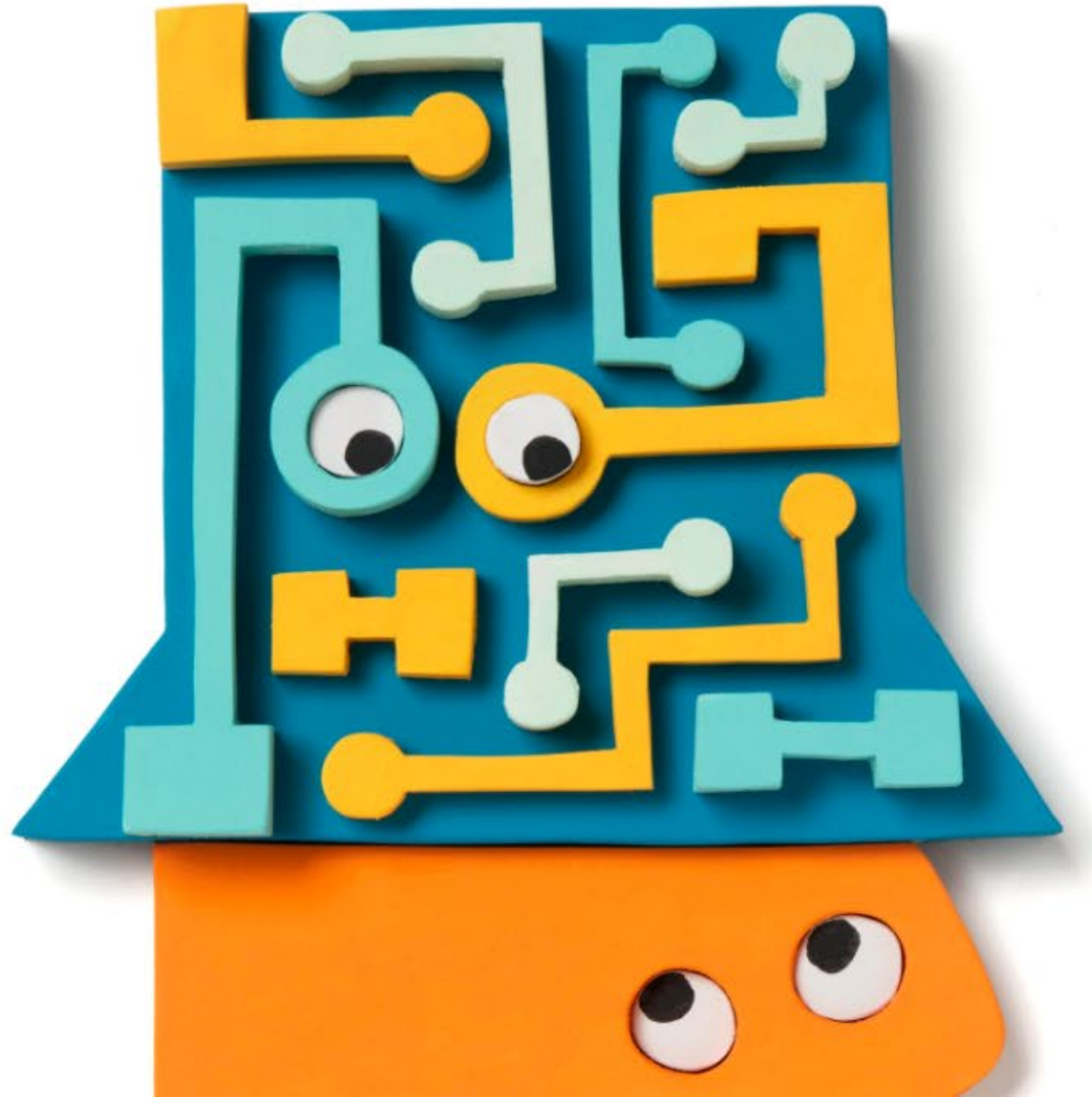
https://www.nytimes.com/2020/04/08/technology/ai-creative-software-language.html

The New York Times

Account

You Can't Spell Creative Without A.I.

Advances in software applications that process human language lie at the heart of the debate over whether computer technologies will enhance or even substitute for human creativity.



“There has long been a dream of using A.I. to help with scientific discovery, and now the question is, can we do that?”

-- Oren Etzioni, the chief executive of the Allen Institute for Artificial Intelligence

COVID-19 Open Research Dataset Challenge (hosted by Kaggle)

What is it?

A series of important questions designed to inspire the community to use CORD-19 to find new insights about the COVID-19 pandemic including the

natural history, transmission, and diagnostics for the virus,

- ✓ management measures at the human-animal interface,
- ✓ lessons from previous epidemiological studies,
- ✓ and more.

NIST and OSTP Launch Effort to Improve Search Engines for COVID-19 Research

April 15, 2020

"…. The TREC-COVID program goals include creating datasets and using an independent assessment process that will help search engine developers to evaluate and optimize their systems in meeting the needs of the research and health-care communities."

https://www.nist.gov/news-events/news/2020/04/nist-and-ostp-launch-effort-improve-search-engines-covid-19-research

Rensselaer
why not change the world?®

Drug re-repurposing analysis [*still early stage*]

ASReview

Tool for researchers and medical doctors to facilitate the reading of literature on the coronavirus.

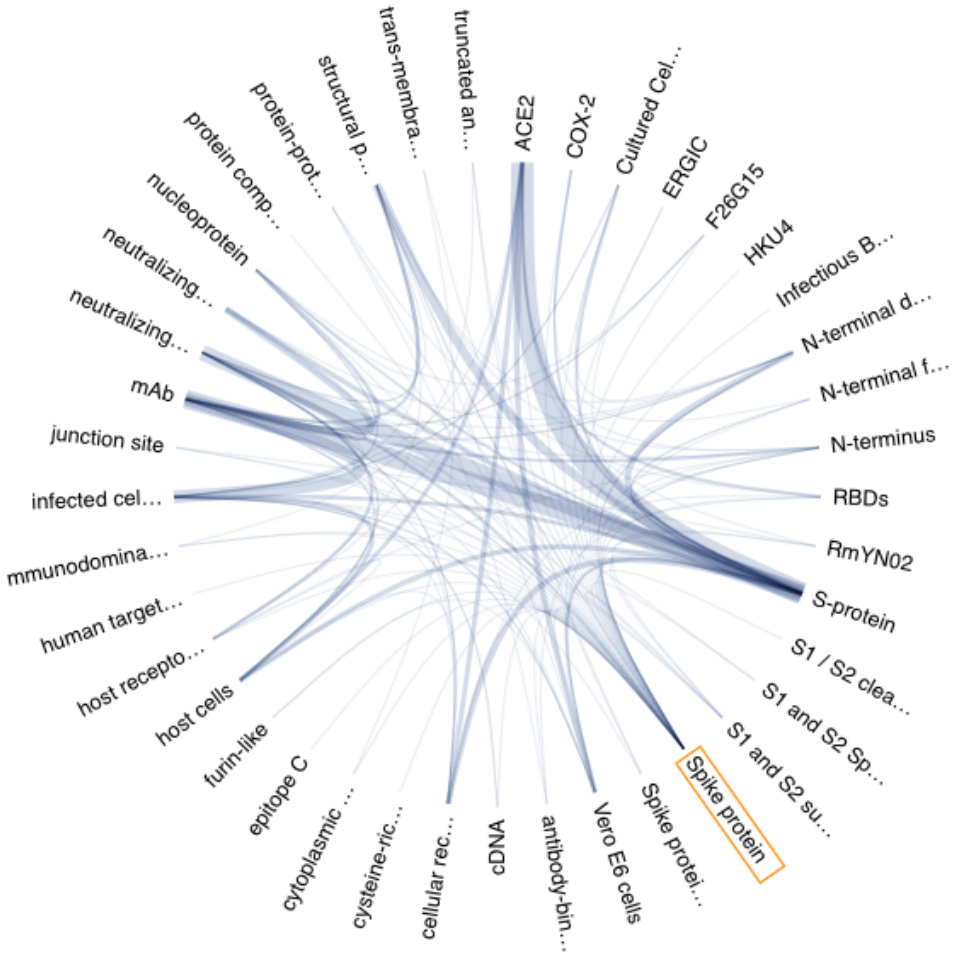
MAYO CLINIC

JOHNS HOPKINS UNIVERSITY

Semantic annotations [*still early stage*]

CoViz

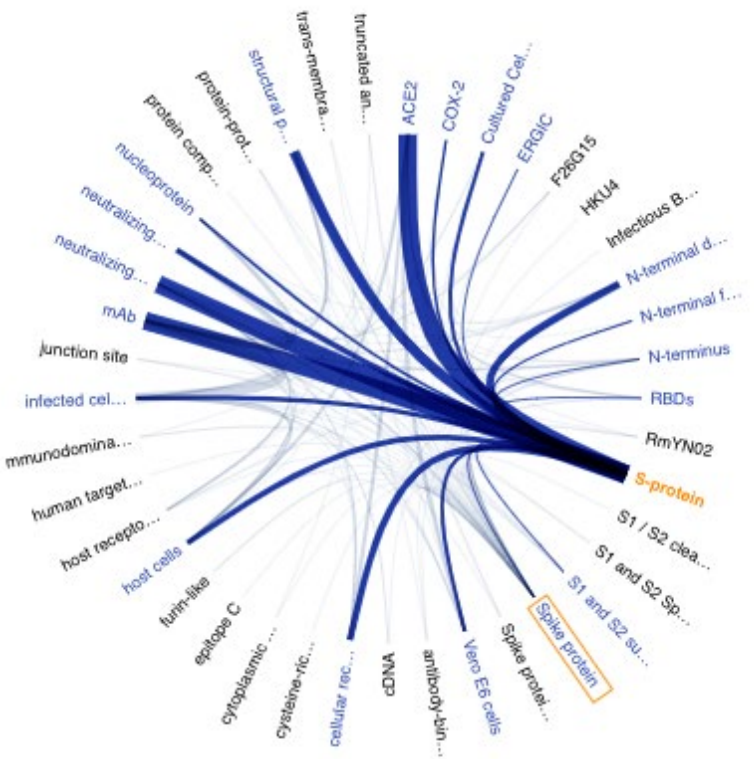
The network of top proteins/genes/cells associated with "Spike protein" in the corpus



Click on an edge above to see associated papers

https://coviz.apps.allenai.org/jnlpba/

The network of top proteins/genes/cells associated with "Spike protein" in the corpus



Papers with co-references to "S-protein" and "ACE2"

A Multiscale and Comparative Model for Receptor Binding of 2019 Novel Coronavirus and the Implication of its Life Cycle in Host Cells

Zhaoqian Su, Yinghao Wu

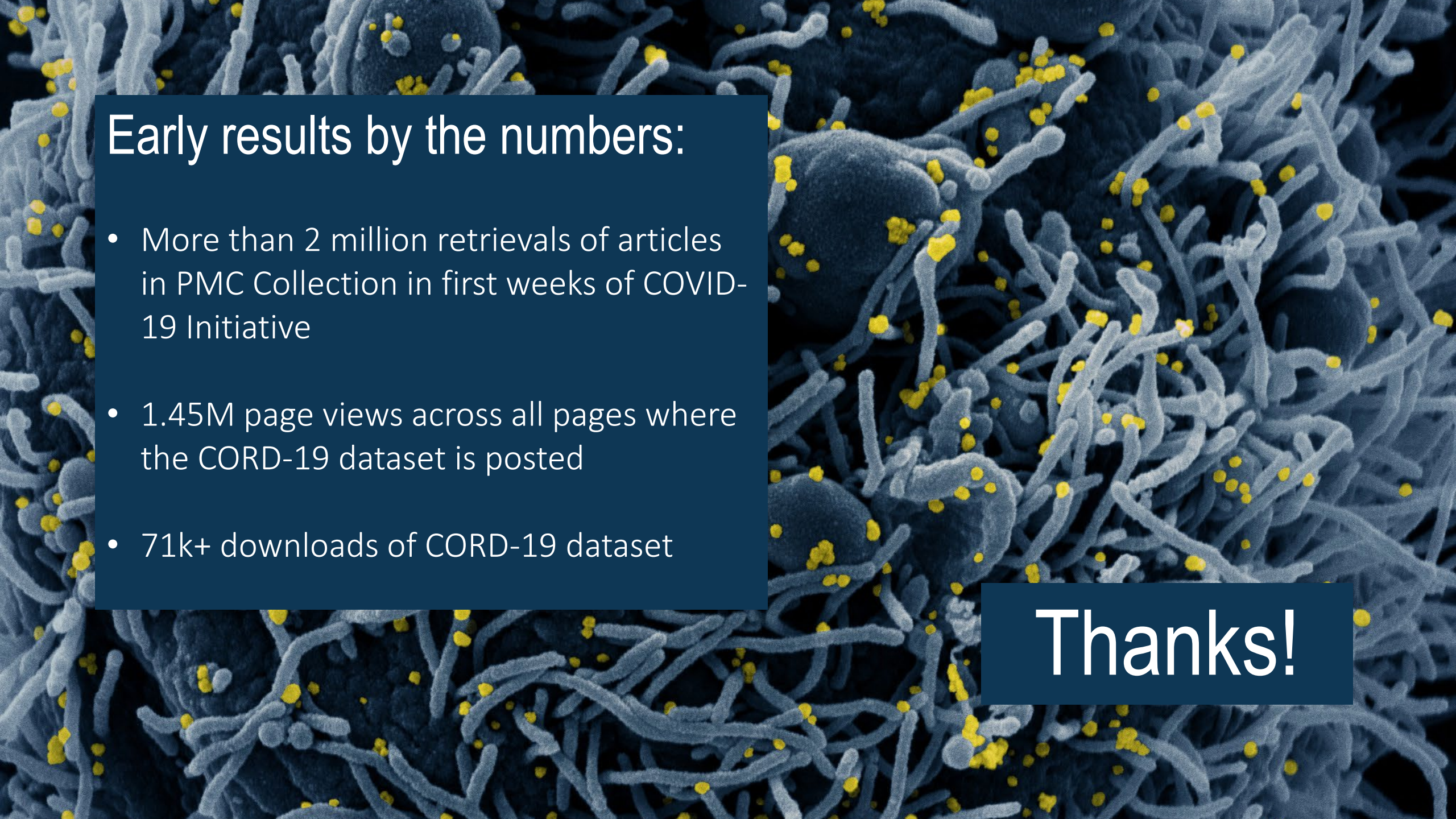
biorxiv • 2020-02-21

The respiratory syndrome caused by a new type of coronavirus has been emerging from 2 China and caused more than 1000 death globally since December 2019. This new virus, 3 called 2019 novel coronavirus (2019-nCoV) uses the same receptor called... more

The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2

biorxiv • 2020-03-17

The rapid and escalating spread of SARS coronavirus 2 (SARS-CoV-2) 6 poses an immediate public health emergency, and no approved therapeutics or 7 vaccines are currently available. The viral spike protein S binds ACE2 on host cells to 8 initiate... more



Early results by the numbers:

- More than 2 million retrievals of articles in PMC Collection in first weeks of COVID-19 Initiative
- 1.45M page views across all pages where the CORD-19 dataset is posted
- 71k+ downloads of CORD-19 dataset

Thanks!