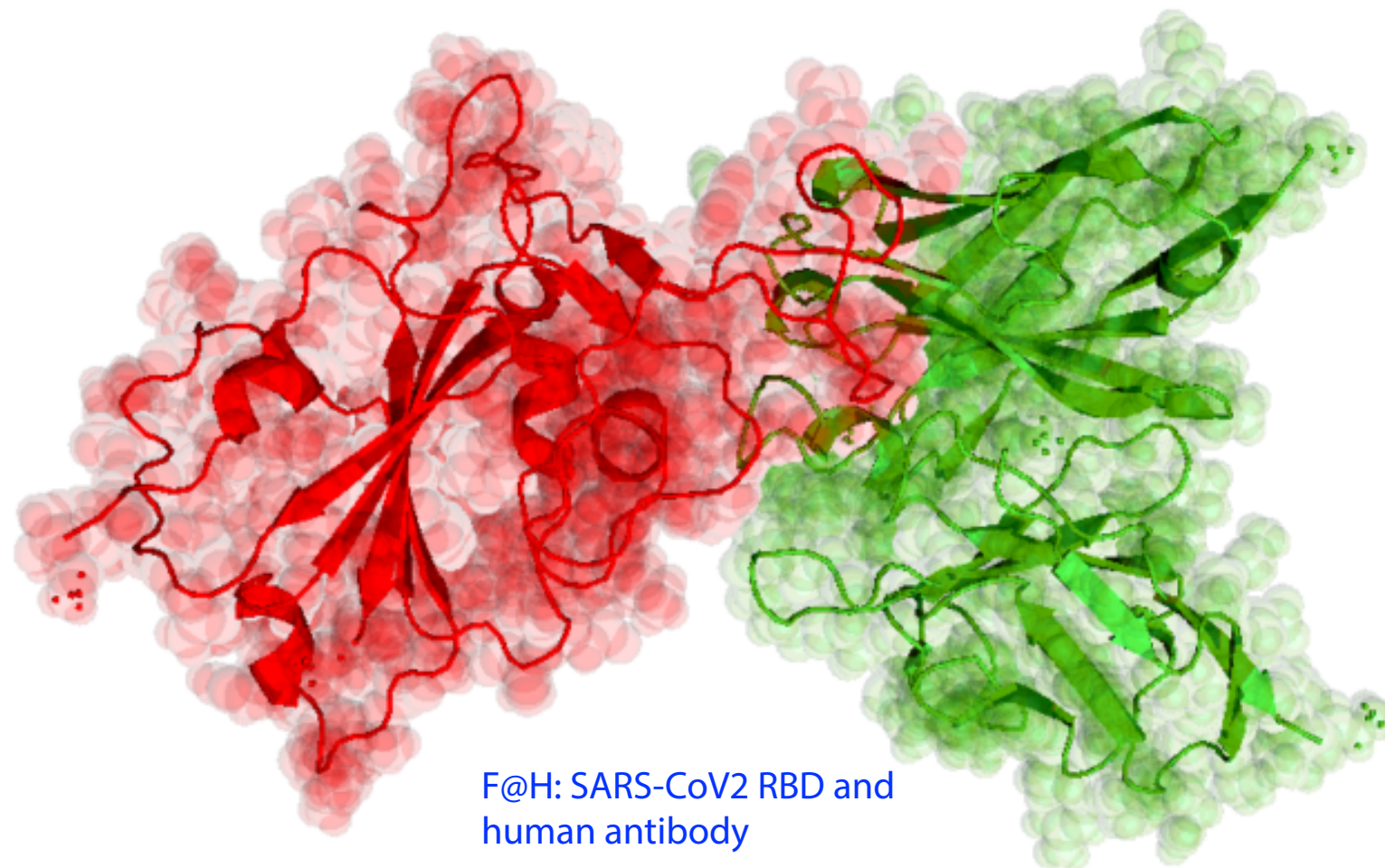


Info about Folding@Home and related Activities

by Peter Elmer, [Markus Elsing](#)

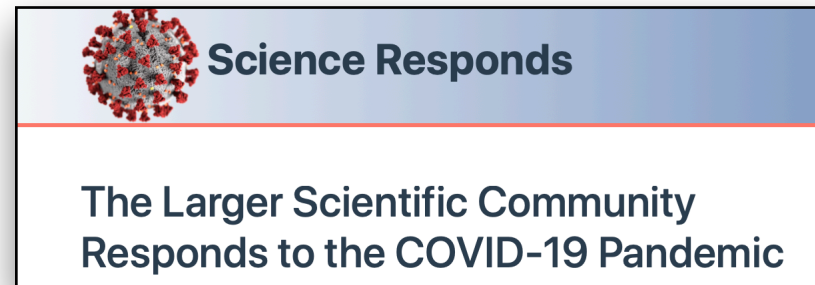
for all people contributing to our community efforts
science-responds.org



F@H: SARS-CoV2 RBD and
human antibody

Folding@Home

- one of the first citizen science initiatives picked up by HEP
 - ➔ a number of parallel initiatives (Jan v. Eldik, Alessandra, Lukas+Mario, ...)
 - ➔ CERN IT, ATLAS, GridPPP, ... devoted resources to it
 - ➔ our experience is that GPU resources are best suited to run the F@H client
 - ➔ more on technical details and operational results in the talk on "Experience" next
- Peter organised meeting with Folding@Home people
 - ➔ last Monday within the open daily [Science-Responds.org](https://www.science-responds.org) townhall meeting



- Greg Bowman (director of F@H) gave us a brief overview
 - ➔ science team of 3 PIs with their groups
 - ➔ project is expanding fast and my impression was they struggle to keep up

Folding@Home

- number of technical discussion items came up
- horizontal scaling of F@H
 - ➔ they run workload servers and need to scale
 - ➔ apparently on public cloud resources, we discussed helping e.g. by hosting those on WLCG or OSG resources
- defining new workloads to keep clients busy
 - ➔ involves adaptive sampling analysis of data produced
 - ➔ needs high memory machines ($\sim > 256$ GB, dual core, GPU)
- Greg explicitly mentioned the issue how to share their data
 - ➔ see ZENODO discussion, he was very interested in this (see later)
- connection points how we as a community can help...



Rosetta@Home and other Initiatives

- Rosetta@Home is another citizen science project
 - ➔ based on BOINC
- ongoing activities within HEP/WLCG community
 - ➔ Thomas Hartman (DESY), Thorsten Harenberg (Wuppertal) and others packaged it to run on GRID resources
 - ➔ more in the talk about "Experience" after this
- direct contact to Rosetta@Home e.g. via Beate and MPI in Germany to Stanford
 - ➔ send a list of question (next slide)



The Questions...

(https://docs.google.com/document/d/1KKXzu0uU1nV_hLETMzLJja3n9z7cHFIUwOASzoa3bsw/edit)

➔ Beate, James, et al.

Information requested from those seeking collaboration w.r.t. Covid-19

What support or collaboration do you envisage?

How can we help, and with what timeline do you envisage this collaboration to happen?

- (1) Large-scale compute facilities, e.g. you have an application ready to be run, and would like to be able to run it on our facilities and retrieve the results.
 - (a) Requires standard CPU (e.g. x86 compatible) architectures
 - (b) Requires or can benefit from GPU acceleration
 - (c) Requires a special architecture
 - (d) Requires large volumes of input data (see item 2 below)
 - (e) Requires parallelization (OpenMPI, Dask, Spark, etc.)
 - (f) Requires outside network access
 - (g) Order-of-magnitude estimate of the total wall clock time required
- (2) Storage and organisation of data, and/or web hosting or other web services
 - (a) Order-of-magnitude estimate of the data volumes
 - (b) Type of data to be stored, and especially, are there confidentiality or other restrictions concerning its use
 - (c) Required level of access to the data once stored
 - (d) Web based service with possibility of structured data insertion to ease analysis
 - (e) Making data publicly available
 - (f) Integration with 3rd-party data stores
- (3) Expert assistance
 - (a) Optimisation of data formats
 - (b) Optimisation of software
 - (c) Assistance with data analysis, including statistics/ML expertise and advice on relevant tools/environments
 - (d) Computing infrastructure assistance (deployment, configuration, technology selection)
 - (e) Form teams to tackle specific problems / topics in a coordinated way (e.g. as follow-up to, e.g., the [versusvirus.ch](https://www.virusshare.ch/) hackathon) and on the medium-term
- (4) Setting up communication channels
 - (a) Sharing information within your community
 - (b) Chat like infrastructure
 - (c) Help setting up an instance of our meeting agenda management system (Indico)



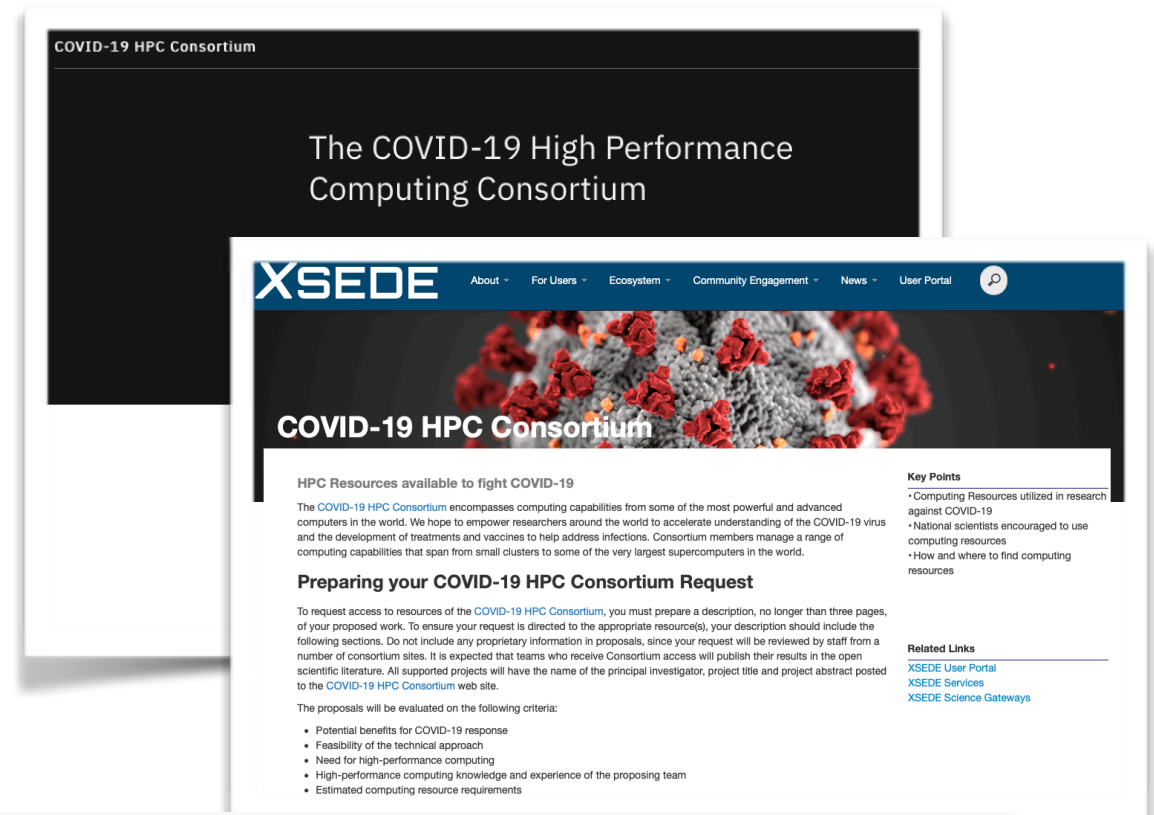
HPC Consortia against COVID 19

➔ US HPC initiative

<https://www.ibm.com/covid19/hpc-consortium>

➔ uses XSEDE portal to collect project proposals

<https://www.xsede.org/covid19-hpc-consortium>



➔ PRACE initiative in EU

<https://prace-ri.eu/prace-support-to-mitigate-impact-of-covid-19-pandemic/>



➔ shall we as WLCG consider to join ? And/or make use of our WLCG science network ?



ZENODO as the Portal for Open Data

- we definitely serve a need of the medical science community
 - ➔ see F@H discussion, CERN contacts with EMBL

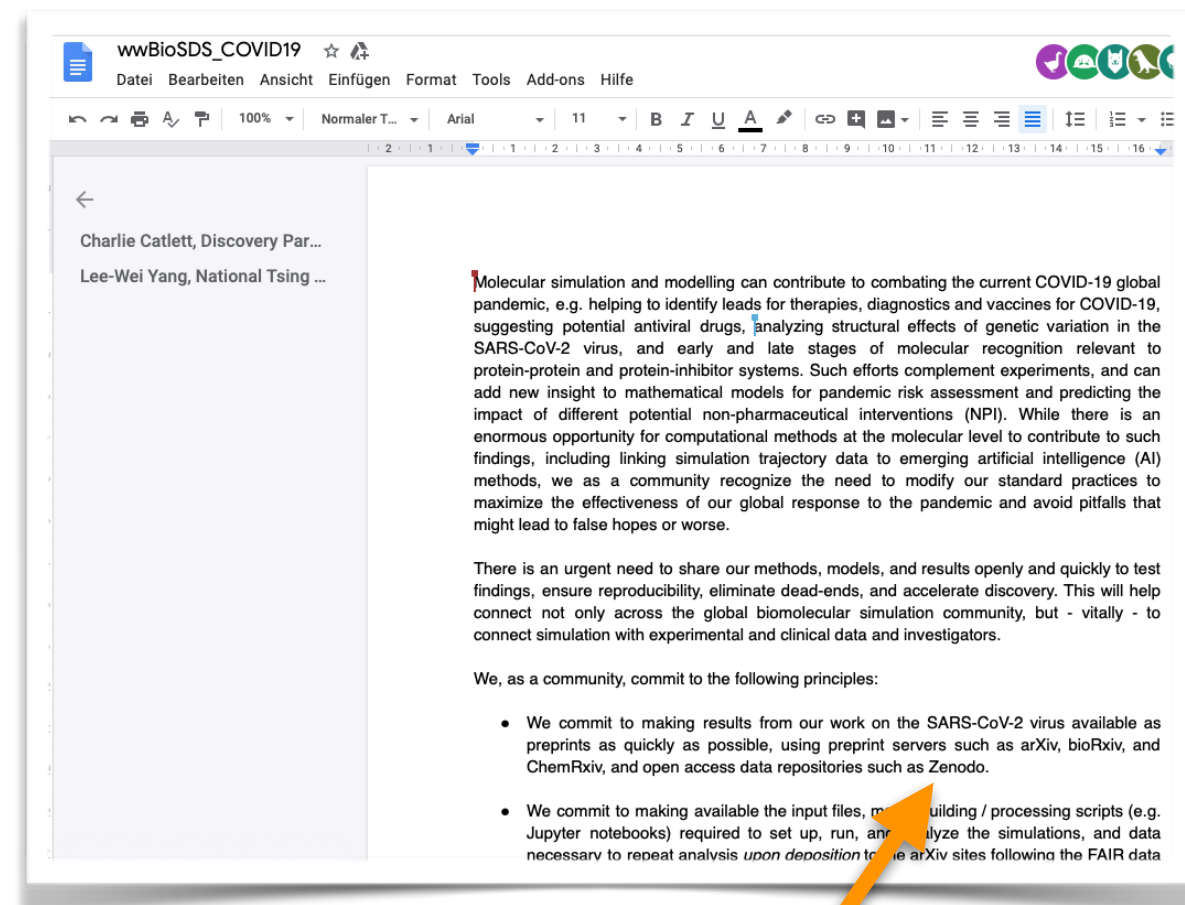
- US to bio-med simulation consortium MoISSI

- ➔ initiative for open data and information sharing: [wwBioSDS_COVID19](#)
- ➔ goes viral - already 5 pages of signatures from Bio-Med community
- ➔ to be published in Springer/Nature (?)

<https://docs.google.com/document/d/11LXHdxmRkLjPaybM9fya9yaSQhj1xZMONQXnIZDK7BI/edit>

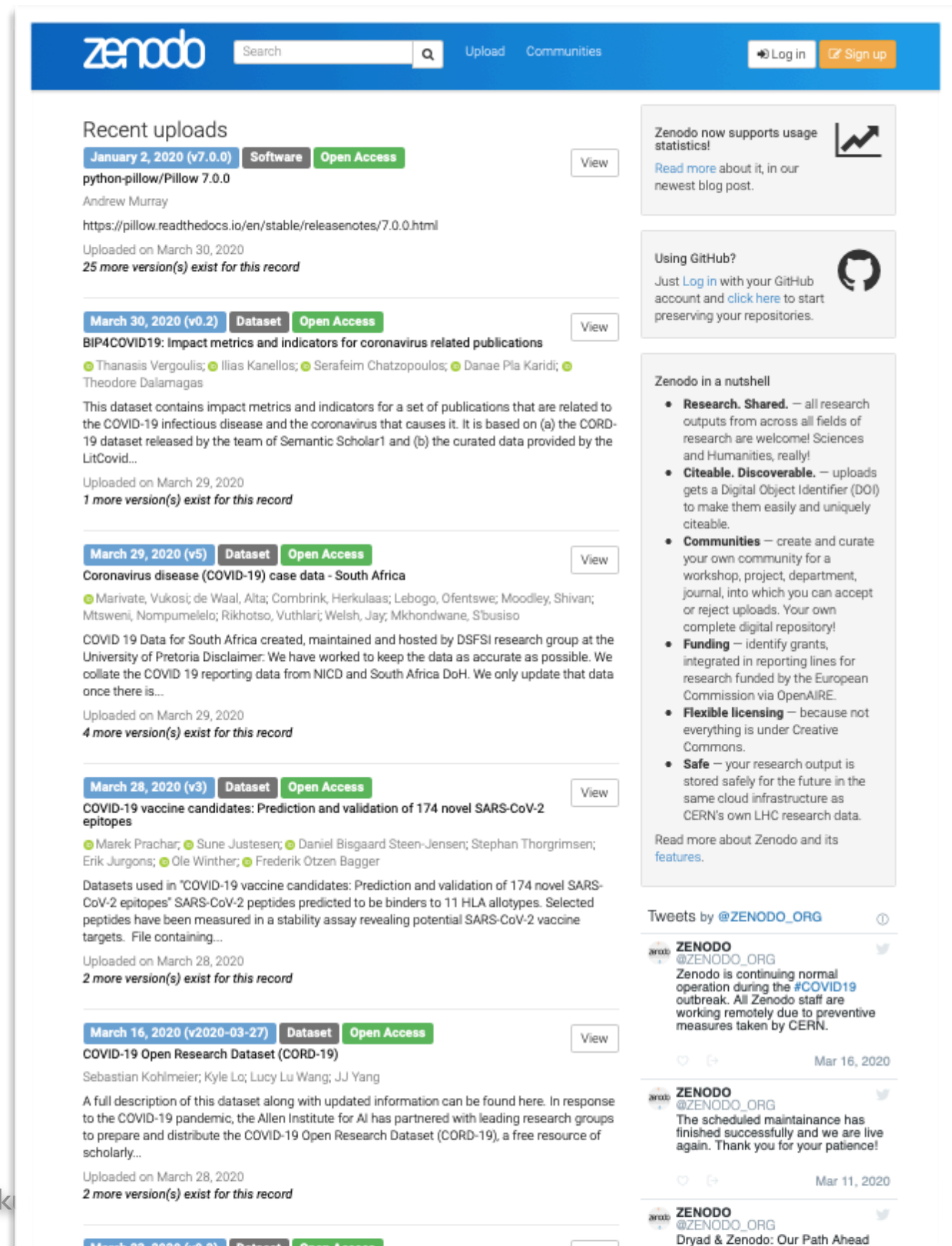
- mentions ZENODO !!!!!

- ➔ need to be prepared
- ➔ sharing large files effectively, ...
- ➔ 50 GB limit already lifted



ZENODO as the Portal for Open Data

- almost entire page 1 of ZENODO is COVID-19 related



The screenshot displays the Zenodo website interface. At the top, there is a navigation bar with the Zenodo logo, a search bar, and links for 'Upload', 'Communities', 'Log in', and 'Sign up'. The main content area is titled 'Recent uploads' and lists several datasets, all of which are COVID-19 related. Each entry includes the upload date, version number, dataset type, and 'Open Access' status. The datasets listed are:

- January 2, 2020 (v7.0.0) Software Open Access**: python-pillow/Pillow 7.0.0 by Andrew Murray. URL: <https://pillow.readthedocs.io/en/stable/releasenotes/7.0.0.html>. Uploaded on March 30, 2020. 25 more version(s) exist for this record.
- March 30, 2020 (v0.2) Dataset Open Access**: BIP4COVID19: Impact metrics and indicators for coronavirus related publications by Thanasis Vergoulis, Ilias Kanellos, Serafeim Chatzopoulos, Danae Pla Karidi, and Theodore Dalamagas. Uploaded on March 29, 2020. 1 more version(s) exist for this record.
- March 29, 2020 (v5) Dataset Open Access**: Coronavirus disease (COVID-19) case data - South Africa by Marivate, Vukosij, de Waal, Alta, Combrink, Herkulaas, Lebogo, Ofentswe, Moodley, Shivan, Mtsweni, Nompumelelo, Rikhotso, Vuthlari, Welsh, Jay, Mkhondwane, S'busiso. Uploaded on March 29, 2020. 4 more version(s) exist for this record.
- March 28, 2020 (v3) Dataset Open Access**: COVID-19 vaccine candidates: Prediction and validation of 174 novel SARS-CoV-2 epitopes by Marek Prachar, Sune Justesen, Daniel Bisgaard Steen-Jensen, Stephan Thorgirmsen, Erik Jurgons, Ole Winther, and Frederik Otzen Bagger. Uploaded on March 28, 2020. 2 more version(s) exist for this record.
- March 16, 2020 (v2020-03-27) Dataset Open Access**: COVID-19 Open Research Dataset (CORD-19) by Sebastian Kohlmeier, Kyle Lo, Lucy Lu Wang, and JJ Yang. Uploaded on March 28, 2020. 2 more version(s) exist for this record.

On the right side of the page, there are several informational boxes:

- Zenodo now supports usage statistics!** with a 'View' button and a link to 'Read more about it, in our newest blog post.'
- Using GitHub?** with a GitHub icon and a link to 'Just Log in with your GitHub account and click here to start preserving your repositories.'
- Zenodo in a nutshell** with a list of features:
 - Research. Shared.** – all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
 - Citeable. Discoverable.** – uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
 - Communities** – create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
 - Funding** – identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
 - Flexible licensing** – because not everything is under Creative Commons.
 - Safe** – your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.

At the bottom right, there is a section for 'Tweets by @ZENODO_ORG' showing three tweets from the Zenodo organization regarding their operations during the COVID-19 outbreak.

Mark

