Info about Folding@Home and related Activities
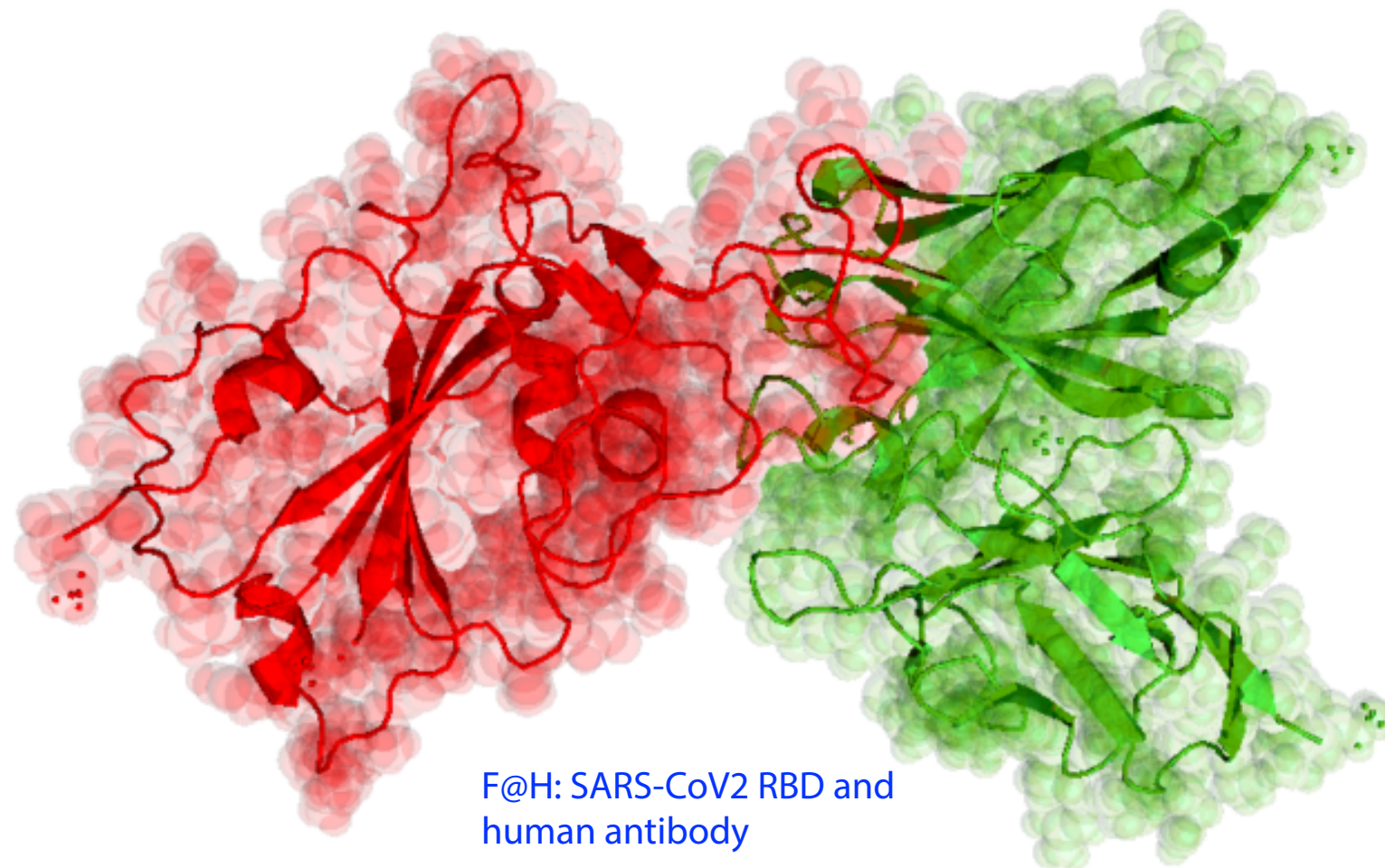
by Peter Elmer, Markus Elsing
for all people contributing to our community efforts
science-responds.org

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 townhall meeting

Science Responds

The Larger Scientific Community
Responds to the COVID-19 Pandemic
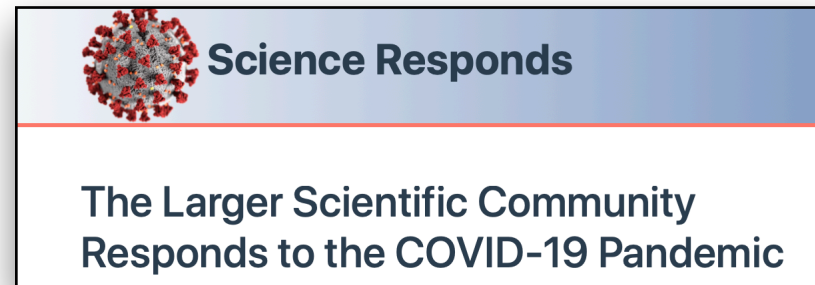
- **Greg Bowman (director of F@H) gave us a brief overview**
  - ➡ science team of 3 PIs with their groups
  - ➡ project is expanding fast and my impression was they struggle to keep up

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 (~>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 Standford
  - ➡ send a list of question (next slide)

The Questions…

➡ 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 hackathon) and on the medium-term
(4) Setting up communication channels
   (a) Sharing information within your community
   (b) Chat like infrastructure
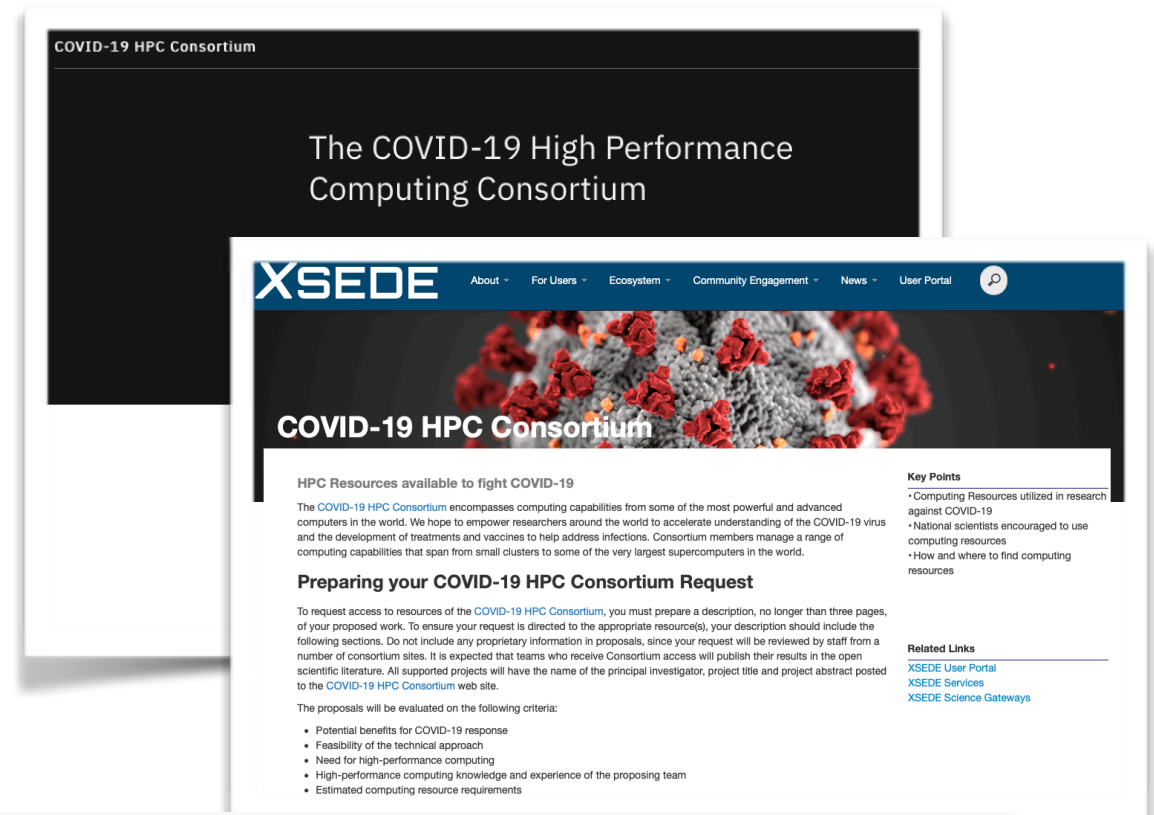   (c) Help setting up an instance of our meeting agenda management system (Indico)

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 definite 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
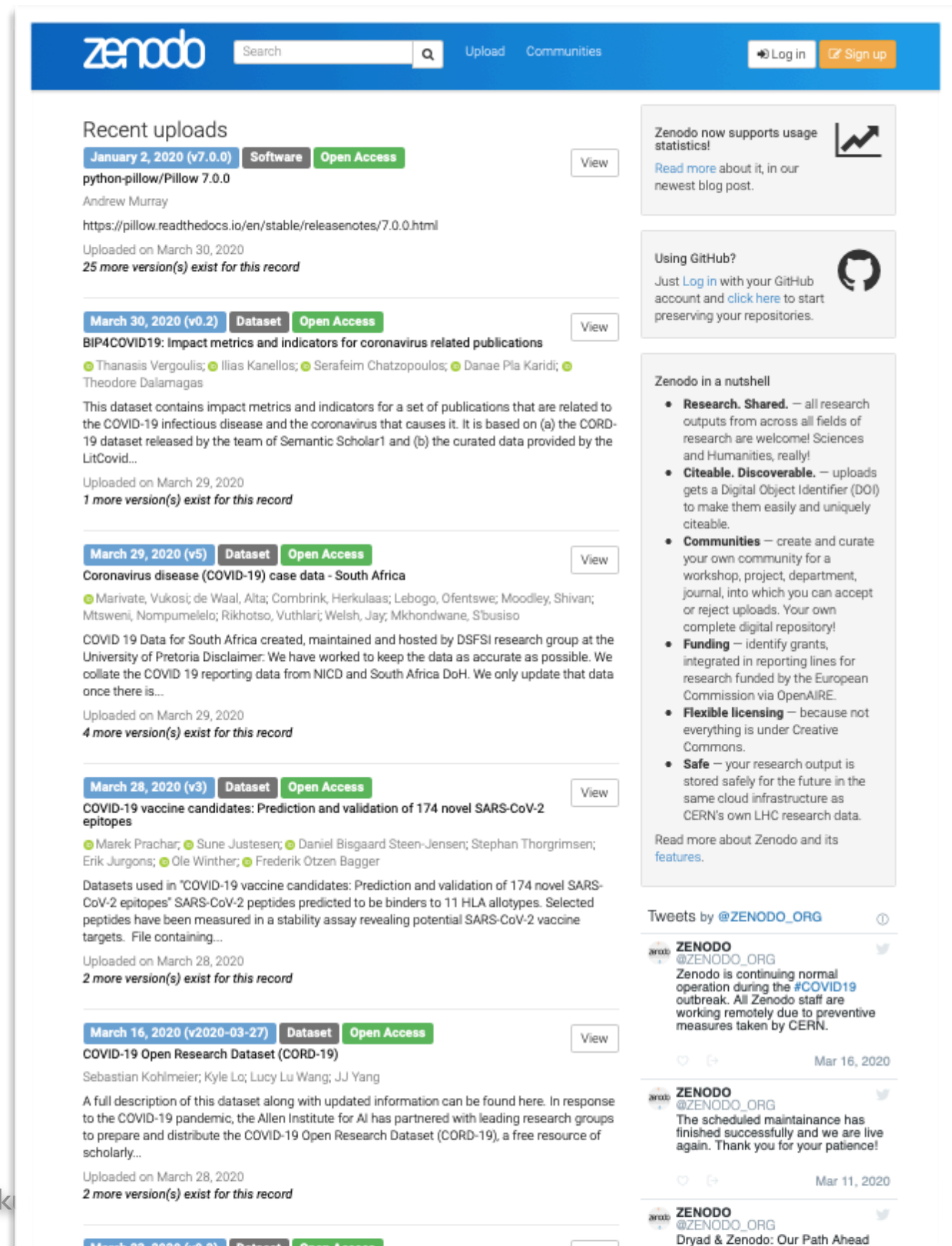
- **mentions ZENODO !!!!!!**
  - ➡ need to be prepared
  - ➡ sharing large files effectively, …
  - ➡ 50 GB limit already lifted
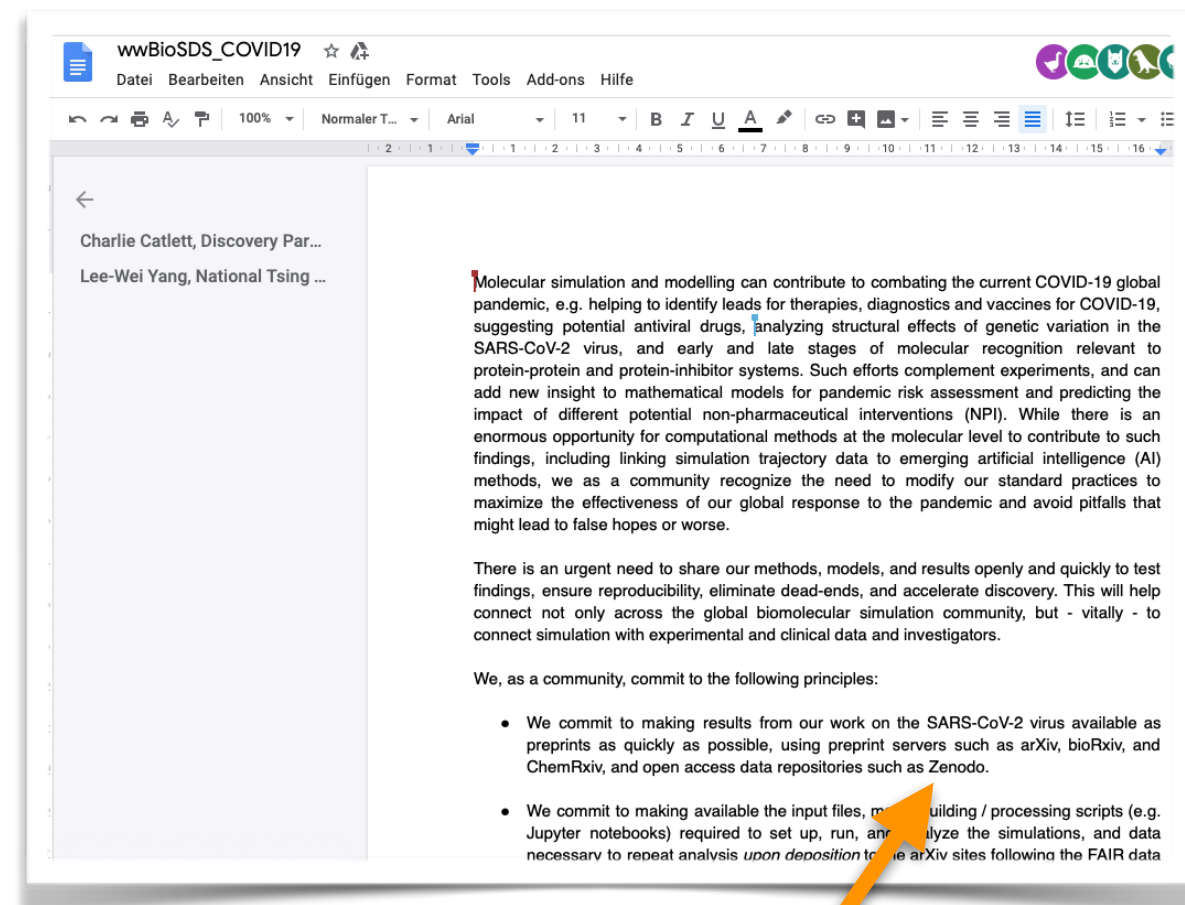
ZENODO as the Portal for Open Data

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