



CERN against COVID-19 Task Force Computing Initiatives

Alberto Di Meglio (IT/DI) on behalf of the many contributors in the Task Force and beyond

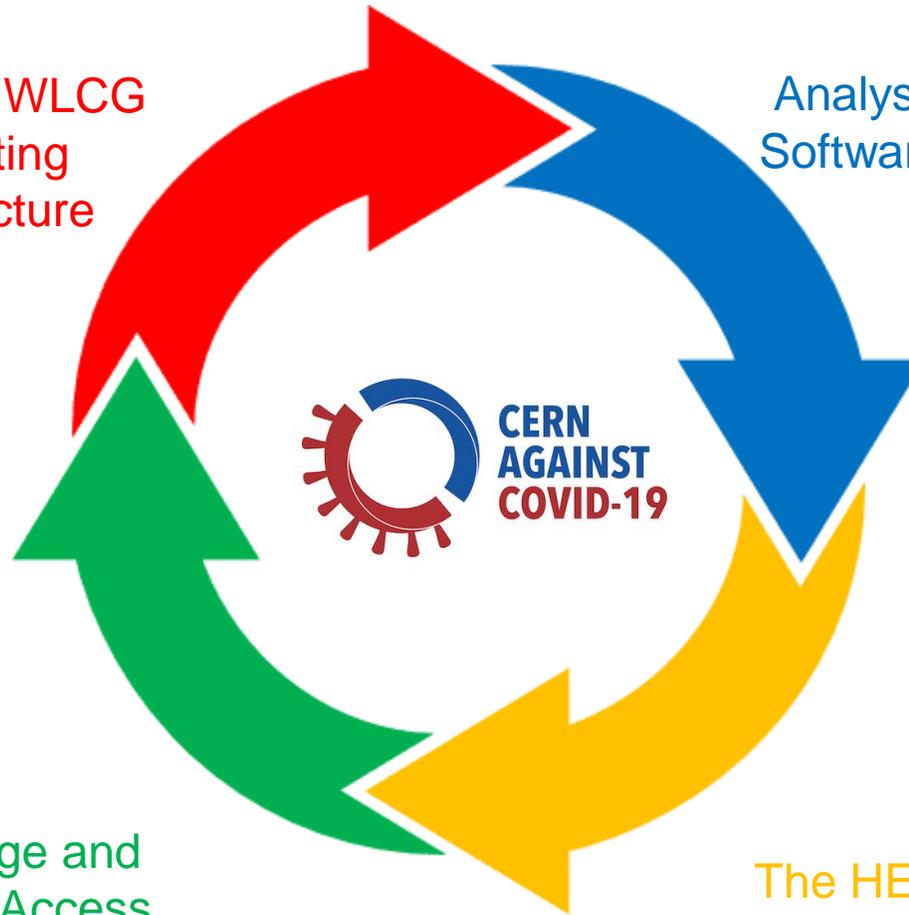
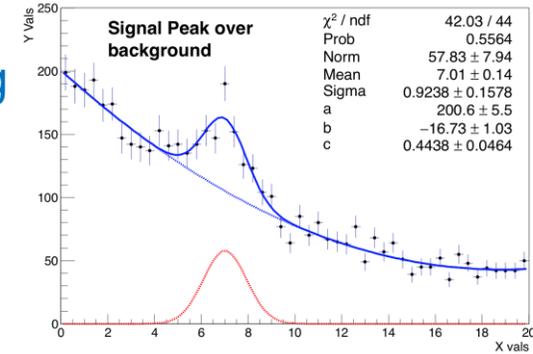
Task Force Meeting – CERN against COVID-19 (20/05/2020)

Areas of Computing Initiatives



CERN and WLCG
Computing
Infrastructure

Analysis, Simulation
Software Engineering



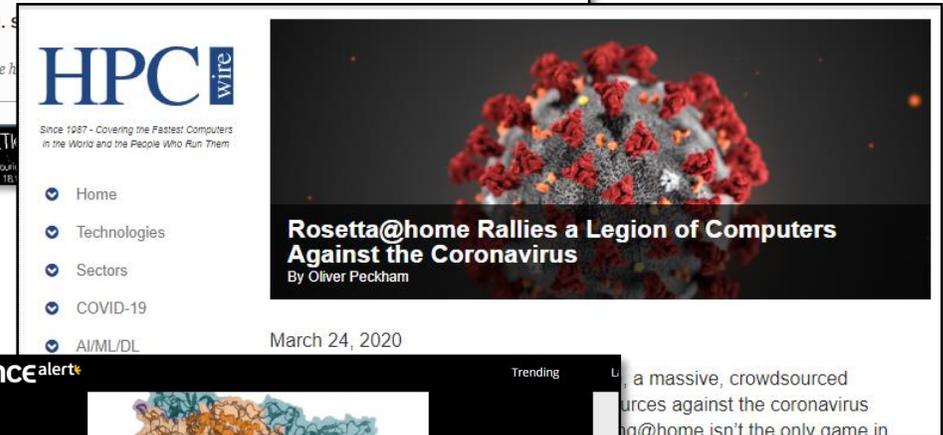
Data Storage and
Open Data Access

The HEP Community
Skills and Creativity



Infrastructure for Volunteer Computing

- Many volunteer computing initiatives have focused on **supporting COVID-19 research**
- Examples include Folding@Home, Rosetta@Home, gene@home, BOINC@TACC, Quarantine@Home, OpenPandemics, or Corona-AI
- CERN is not new to volunteer computing, the **LHC@Home project** (<https://lhathome.web.cern.ch/home>) provides computing resources to LHC experiments and accelerator physics simulations for LHC and HL-LHC from more than **200K** volunteers at an average level of **45 TeraFLOPS**



CERN and WLCG Contributions

- CERN was already contributing to Folding@Home before March as a general support for scientific initiatives
- In March CERN and WLCG (initially mainly from ATLAS) have started a **community effort** to support F@H, which has **rapidly rallied resources** from **all the LHC experiments**. Today CMS and ATLAS are the largest resource providers (Top10 in May worldwide)
- In just two months the community has collectively provided around **1M computing cores** running **millions of jobs per day** from more than **220** volunteer sites and individuals

Donors: CMS and ATLAS
Top10 in May out of more than 1.5M donors

Team:
31st contributor overall
5th in May
out of more than 250k teams worldwide

<https://stats.foldingathome.org/team/38188>

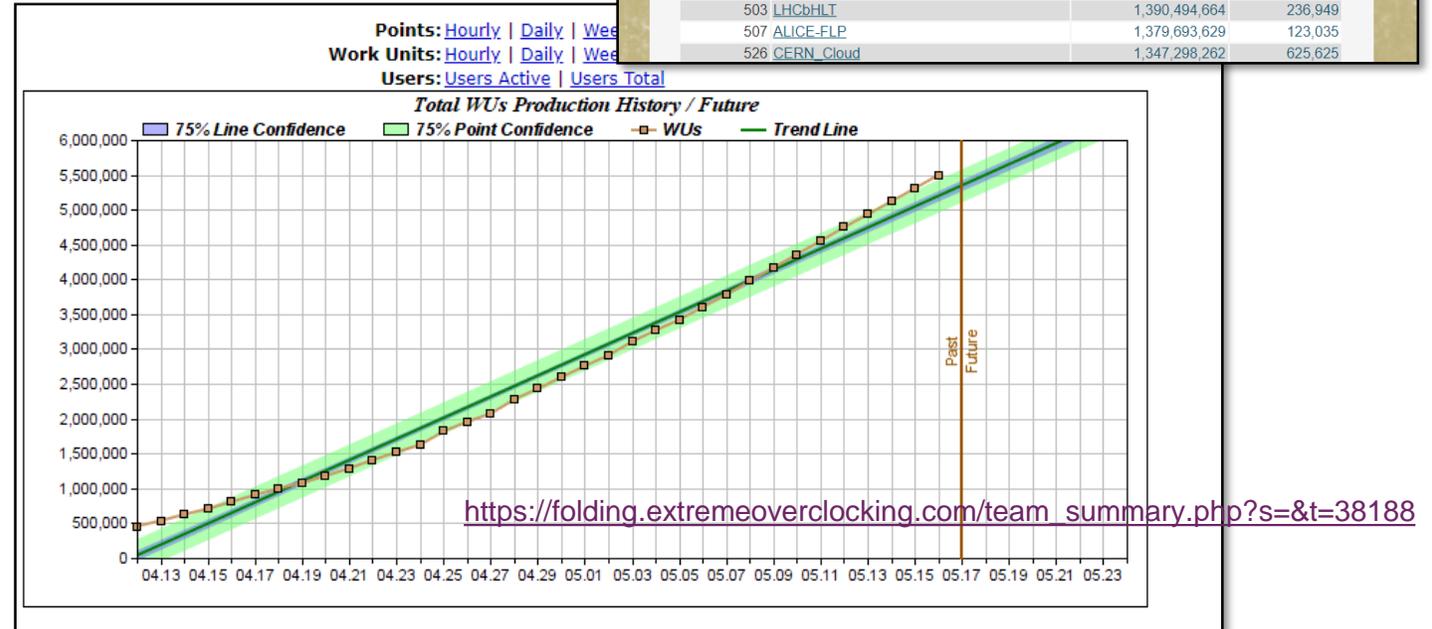


Folding@Home Team: CERN & LHC Computing

Date of last work unit: 2020-05-19 18:56:42
Active CPUs within 60 days: 1,018,282
Team Id: 38188
Grand Score: 20,696,258,327
Work Units: 5,952,658
31 of 253417
Fast Teampage: <http://public.web.cern.ch/public/>
<https://apps.foldingathome.org/teamstats/team38188.html>

Rank Project	Use Nam
1	Anonymous
2	ra9lab
3	NVIDIA_Saturn_V
4	CMS-Experiment
5	ANX_Developers
6	ATLAS_CPU
7	CRT-DC_and_Frontier_Labs
8	sdumont.petrobras
9	Arbutus

Rank	Name	Credit	WUs
53	CMS-Experiment	8,275,282,617	1,711,582
77	ATLAS_CPU	6,579,078,398	1,680,295
503	LHCbHLT	1,390,494,664	236,949
507	ALICE-FLP	1,379,693,629	123,035
526	CERN_Cloud	1,347,298,262	625,625



Data Management Contributions to F@H

- A major consequence of the rapid increase in volunteer resources worldwide was the **need to scale the distribution infrastructure**
- CERN (IT and EP Dep.) **teamed up** with the Bowman Labs (home of F@H), GridPP, STFC and the Hartree HPC centre in the UK, FNAL and UChicago in US and others to **expand the F@H data infrastructure**
- HEP framework and tools such as **RUCIO** and **FTS** (used by the LHC experiments to transfer and manage data worldwide) are being deployed to expand the existing F@H services

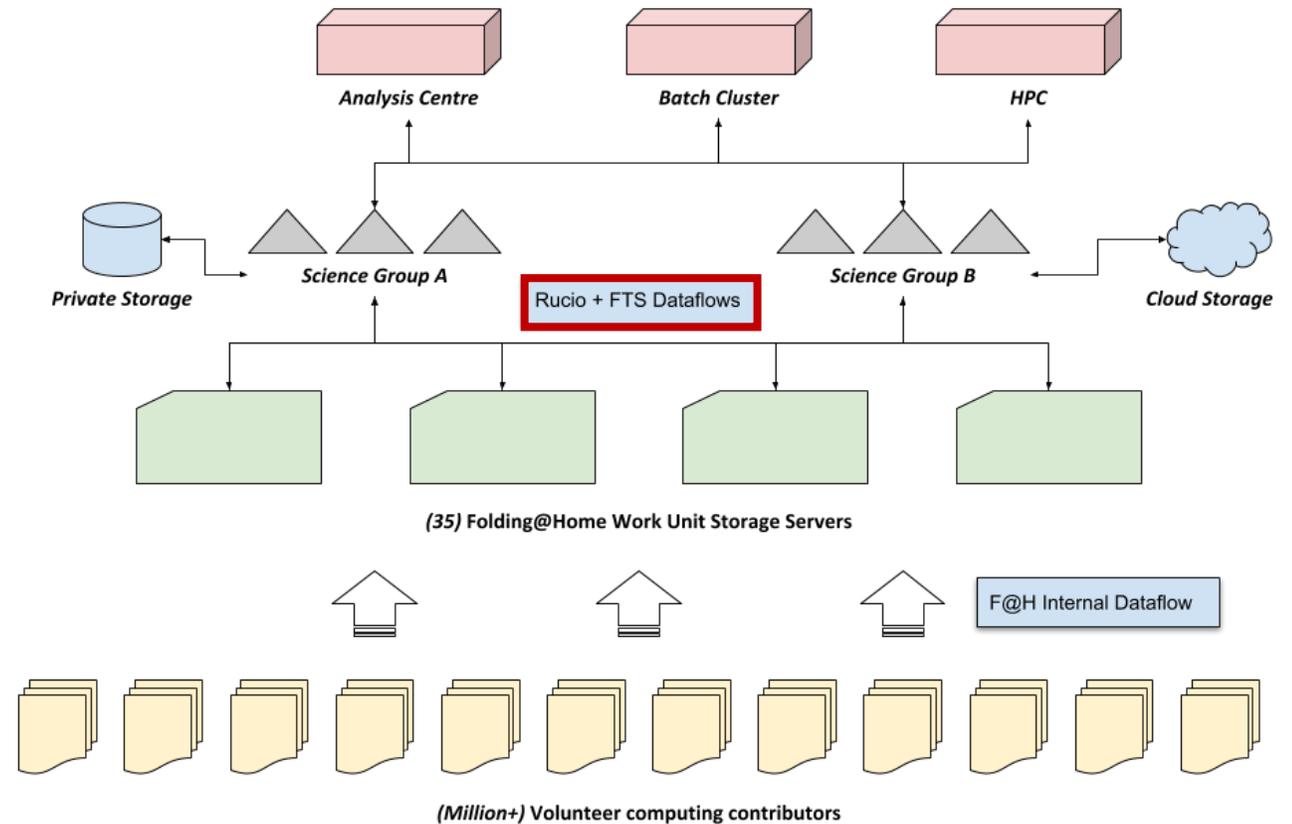
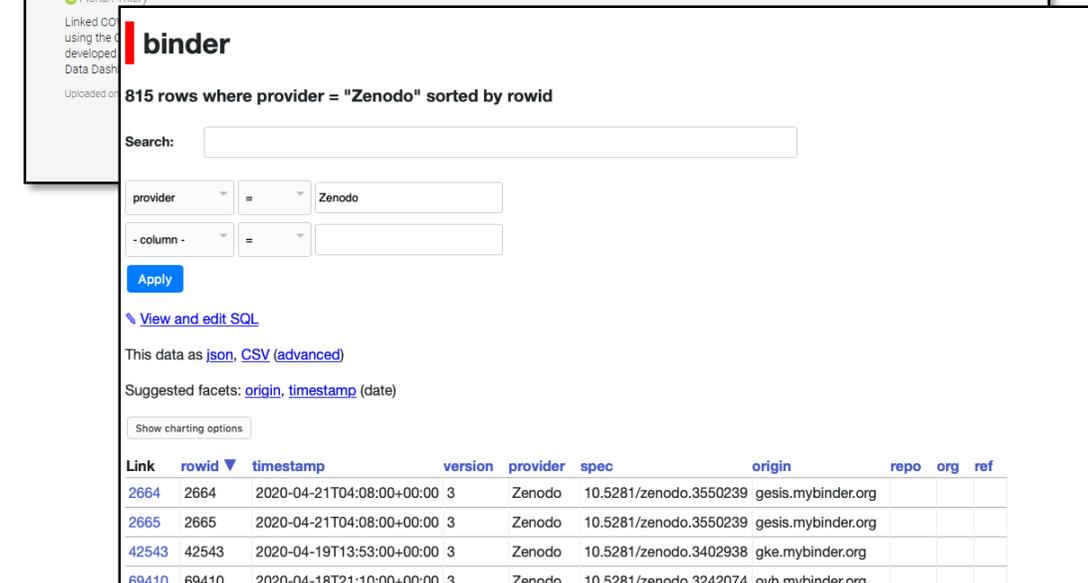
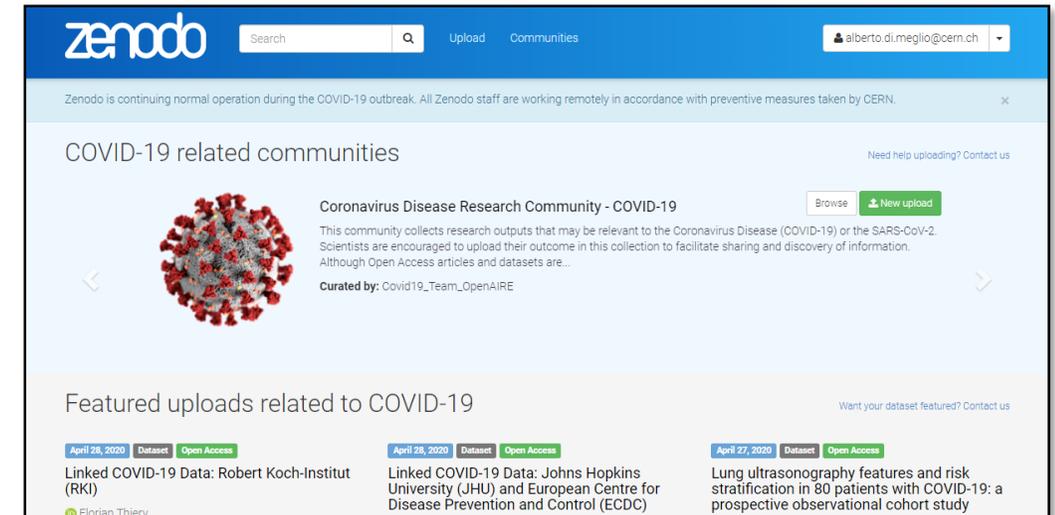


Diagram courtesy of Mario Lassnig

Open Data & Digital Assets Storage

- The worldwide research efforts against COVID-19 have brought to many more people's attention the role of **open data repositories** to support **reproducible, collaborative research**
- CERN was already very active in this area before the current situation with tools like **Zenodo** (<https://zenodo.org>) and the **CERN Open Data portal** (<http://opendata.cern.ch/>)
- Terms of Use for Zenodo have been revised in anticipation of an **increase in requests** to store open access medical research data
- **Integration of Zenodo** within the data flows of many initiatives, e.g. Binder (a popular Notebook execution framework, <https://mybinder.org>) shows **815** different Zenodo-based flows



Examples of Datasets in Zenodo

COVID-19 CT Lung and Infection Segmentation Dataset

April 20, 2020

4,486 views | 6,920 downloads

Indexed in **OpenAIRE**

Medical

Raw diffraction data for structure of SARS-CoV-2 main protease with Z1271660837 (ID: mpro-x1226 / PDB: 5RFB)

March 30, 2020

257 views | 23 downloads

Indexed in **OpenAIRE**

Biological

Linked COVID-19 Data: Johns Hopkins University (JHU) and European Centre for Disease Prevention and Control (ECDC)

May 18, 2020

3,178 views | 645 downloads

Indexed in **OpenAIRE**

Standardization

Crowdsourced air traffic data from The OpenSky Network 2020

May 1, 2020

939 views | 179 downloads

Indexed in **OpenAIRE**

Transportations

Open Access Data Initiatives

Circular Health

An initiative by the One Health Centre of Excellence at UFlorida to support the creation of a large-scale federation of open access data repositories for global health research

CERN contributions: creation of a **Zenodo** community and a **GitHub** repository, support for data collection and aggregation, workflows automation (**REANA**)

EMBL-EBI Data Portal (<https://www.ebi.ac.uk/covid-19>)

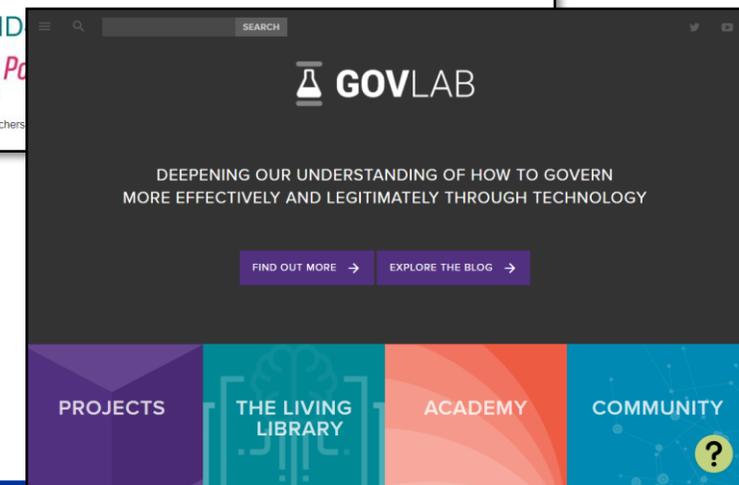
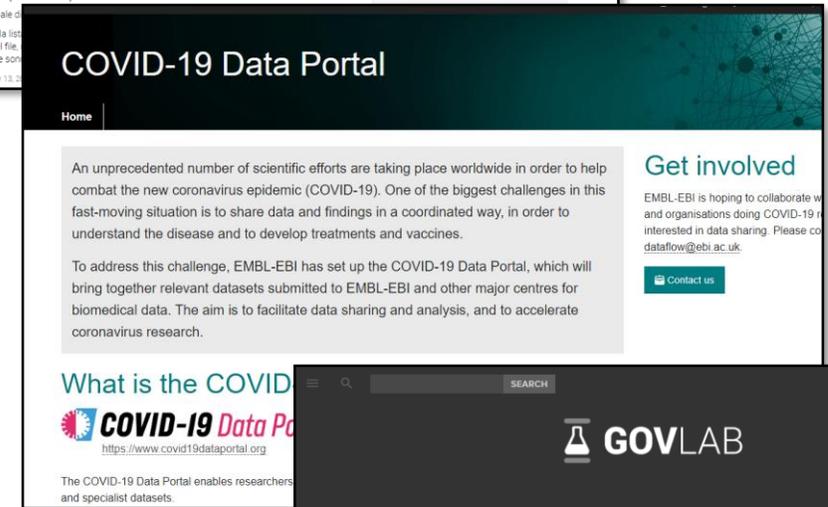
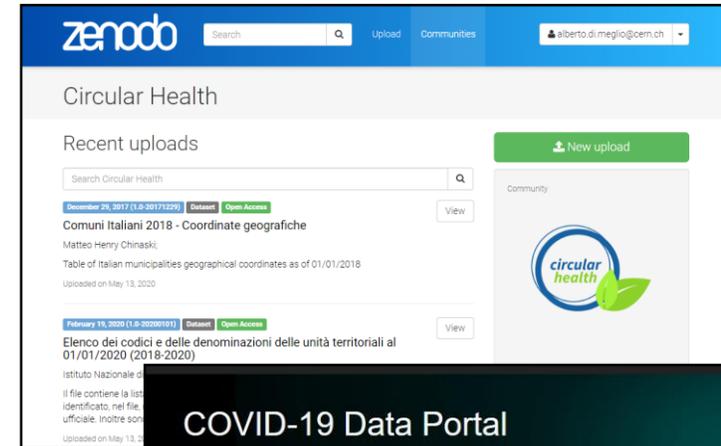
A portal to bring together relevant datasets submitted to EMBL-EBI and other major centres for biomedical data with the aim to facilitate data sharing and analysis, and to accelerate coronavirus research.

CERN contributions: agreements to direct certain types of data to Zenodo, establishment of contacts between the respective data curator teams

The Governance Lab (GovLab, <https://www.thegovlab.org/>)

International initiative to “strengthen the ability of institutions and people to work more openly, collaboratively, effectively and legitimately to make better decisions and solve public problems”

CERN contributions: Initial discussions to share ideas around effective use of open access data and policies



Data Analysis, Simulation, Software Engineering



Health Emergency and Disaster Risk Management

Implementation of the WHO emergency and disaster risk management framework, support for policies, planning, knowledge management

CERN contribution (BE, IT, KT)

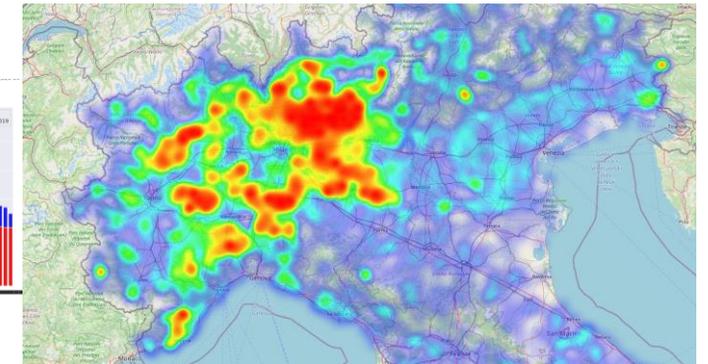
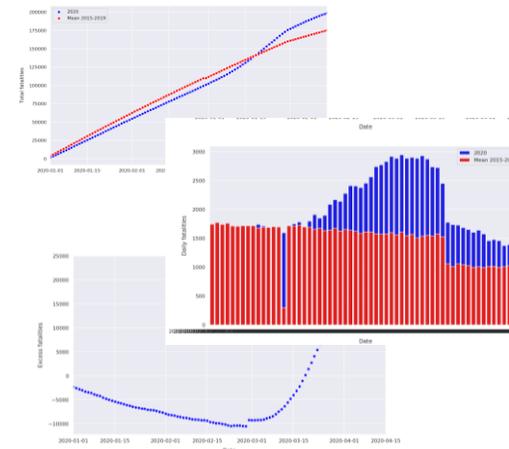
- Collaboration on data architectures optimization of risk analysis and supply chain management
- Possibly part of broader regional initiatives to integrate medical, industrial and governance data systems



Analysis of COVID-19 impact and correlations with health and social factors (co-morbidities, demographics, environmental conditions)

CERN contribution (EP, IT)

- Implementation with Zenodo, REANA, SWAN as part of Circular Health
- Data aggregation, analysis, visualization of public access statistical and demographical data



<https://github.com/CERN/CircularHealth>

Data Analysis, Simulation, Software Engineering

CompBioMed is active in a vast international consortium across Europe and USA working on urgent coronavirus research (UCL, UvA, BSC, SurfSara, IIT, and more)

CERN contribution (EP, IT)

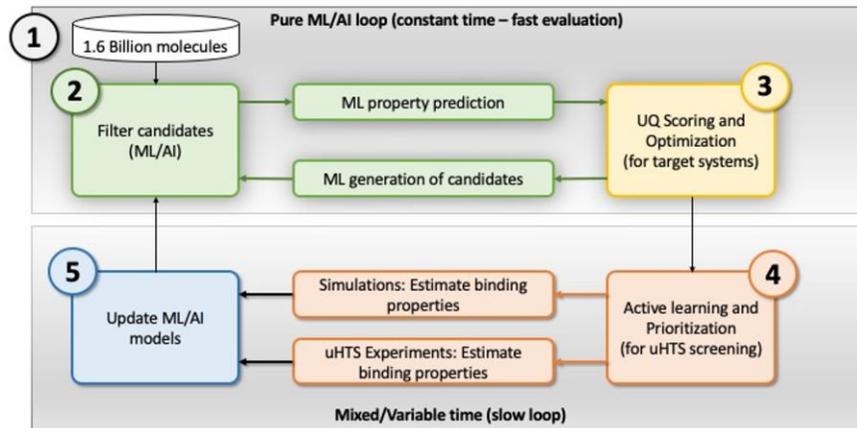
- Expertise in Monte Carlo and GAN simulation applied to molecular dynamics
- Fast trigger/filter algorithms for classification and selection (LHC experiments)

INSTITUTE OF GLOBAL HEALTH

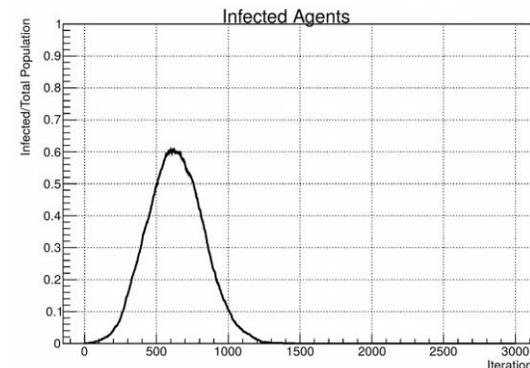
Development of academic global health in collaboration with WHO, UNAIDS, and many other NGOs and foundations

CERN contribution (IT, KT)

- Agent-based simulation of dynamic systems for epidemiologic applications
- Use of BioDynaMo project from CERN openlab/KT
<https://biodynamo.org/>

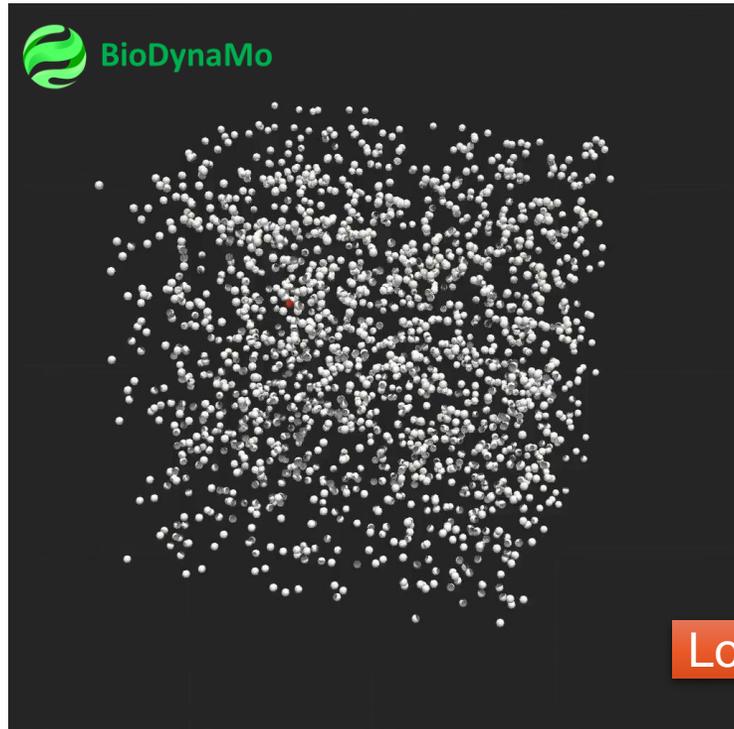


<https://www.compbiomed.eu/coronavirus-computational-drug-screening/>



Examples of Epidemiological Simulations

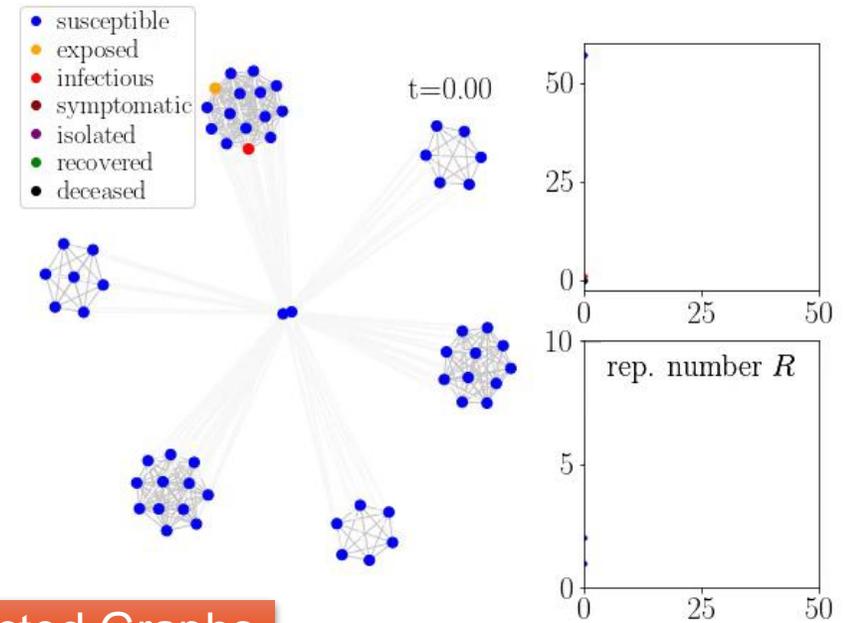
Simulations of virus propagation based on models of infection and recovery rates



Local Agents



Lukas Breitwieser - Ahmad Hesam - Fons Rademakers (IT Dep.)



Connected Graphs

Simulating the Spread of Viruses inside an Organization

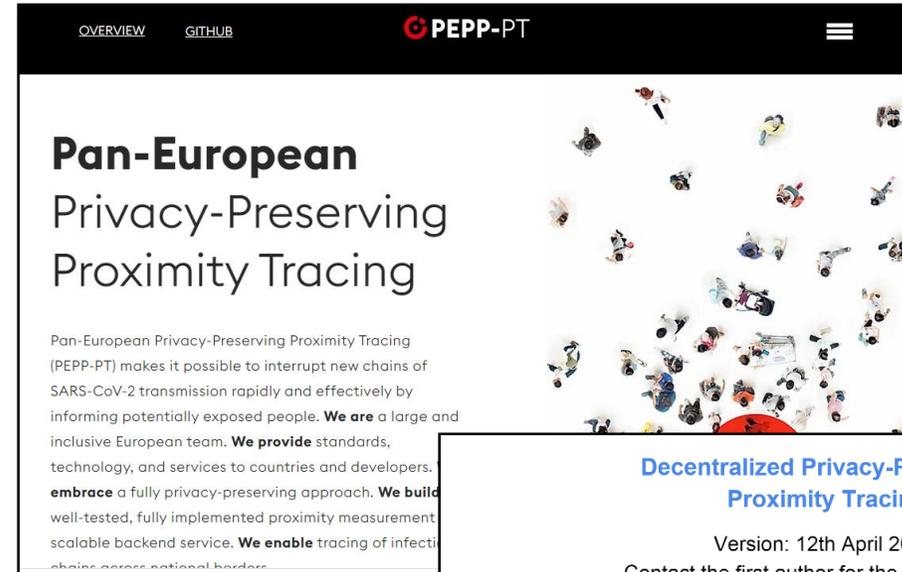
Joachim Kopp (CERN TH Dep. & Uni Mainz)

Contact Tracing Apps

An effort to control COVID-19 propagation and support the return to more flexible living and working conditions, various contact or proximity tracing solutions, ranging from decentralized to centralized, are being considered in many contexts across the world.

Within the wider HEP community (**Science-Responds initiative**), discussions are taking place on **fully open-source solutions and protocols**.

CERN is following discussions from a computing perspective. Of course, any solution should also include legal and privacy analyses.



Discussions with PEPP-PT and the team at EPFL developing the DP-3T protocol are taking place

Decentralized Privacy-Preserving Proximity Tracing

Version: 12th April 2020.
Contact the first author for the latest version.

EPFL: Prof. Carmela Troncoso, Prof. Mathias Payer, Prof. Jean-Pierre Hubaux, Prof. Marcel Salathé, Prof. James Larus, Prof. Edouard Bugnion, Dr. Wouter Lueks, Theresa Stadler, Dr. Apostolos Pyrgelis, Dr. Daniele Antonioli, Ludovic Barman, Sylvain Chatel

ETHZ: Prof. Kenneth Paterson, Prof. Srdjan Čapkun, Prof. David Basin, Dr. Jan Beutel, Dennis Jackson

KU Leuven: Prof. Bart Preneel, Prof. Nigel Smart, Dr. Dave Singelee, Dr. Aysajan Abidin

TU Delft: Prof. Seda Guerses

University College London: Dr. Michael Veale

CISPA: Prof. Cas Cremers

University of Oxford: Dr. Reuben Binns

University of Torino / ISI Foundation Prof. Ciro Cattuto

Community Response

The collective response and engagement of the HEP community went well beyond the projects and activities directly facilitated by the Task Force

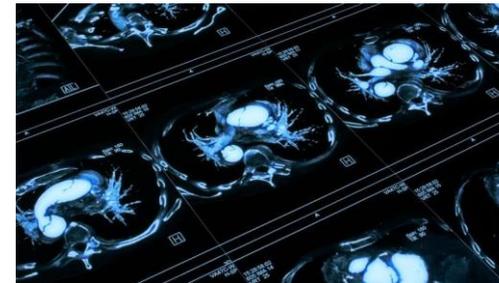


<https://science-responds.org/about/overview/>

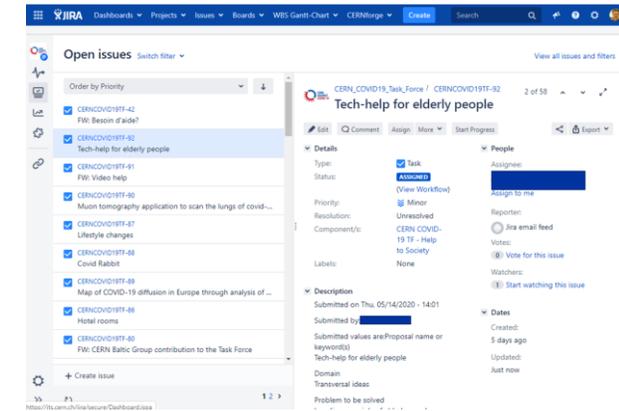
Facilitate interaction between COVID-19 researchers and the broader science community



Helping several health organisations, in order to improve the detection, protection and blocking of potential attacks



Computing projects funded by the KT Medical Apps scheme have adapted their tools to help doctors, e.g. with automated image analysis (CAFEIN) or health monitoring systems (MARCHESE)



Almost 100 different proposals, ideas and offers of help, including computing and data skills for projects and technical support to people in need

Long-Term Impact

- **The relations built with organizations, international initiatives and people have strengthen even more our community**
- **Ways of supporting the initiatives beyond the current emergency are being discussed, for example via project proposals for European and national funds**
 - The collaboration on epidemiologic simulation between CERN and UNIGE in BioDynaMo has just received funds from the EOSC COVID-19 Fast-Track programme to continue the research for the next 9 months
- **Lessons learned are being shared with other organizations**
 - Many of the current activities will be provided as input for the development of services for the European Open Science Cloud (EOSC) and the Health Research & Innovation Cloud (HRIC)
 - Collaborations with other organizations such as WHO or EMBL will continue through existing or new collaboration agreements



home.cern