SciLifeLab

The Swedish COVID-19 data portal: From leveraging infrastructure investment to brokering Public Health submission

Riding the Next Wave of Research Data

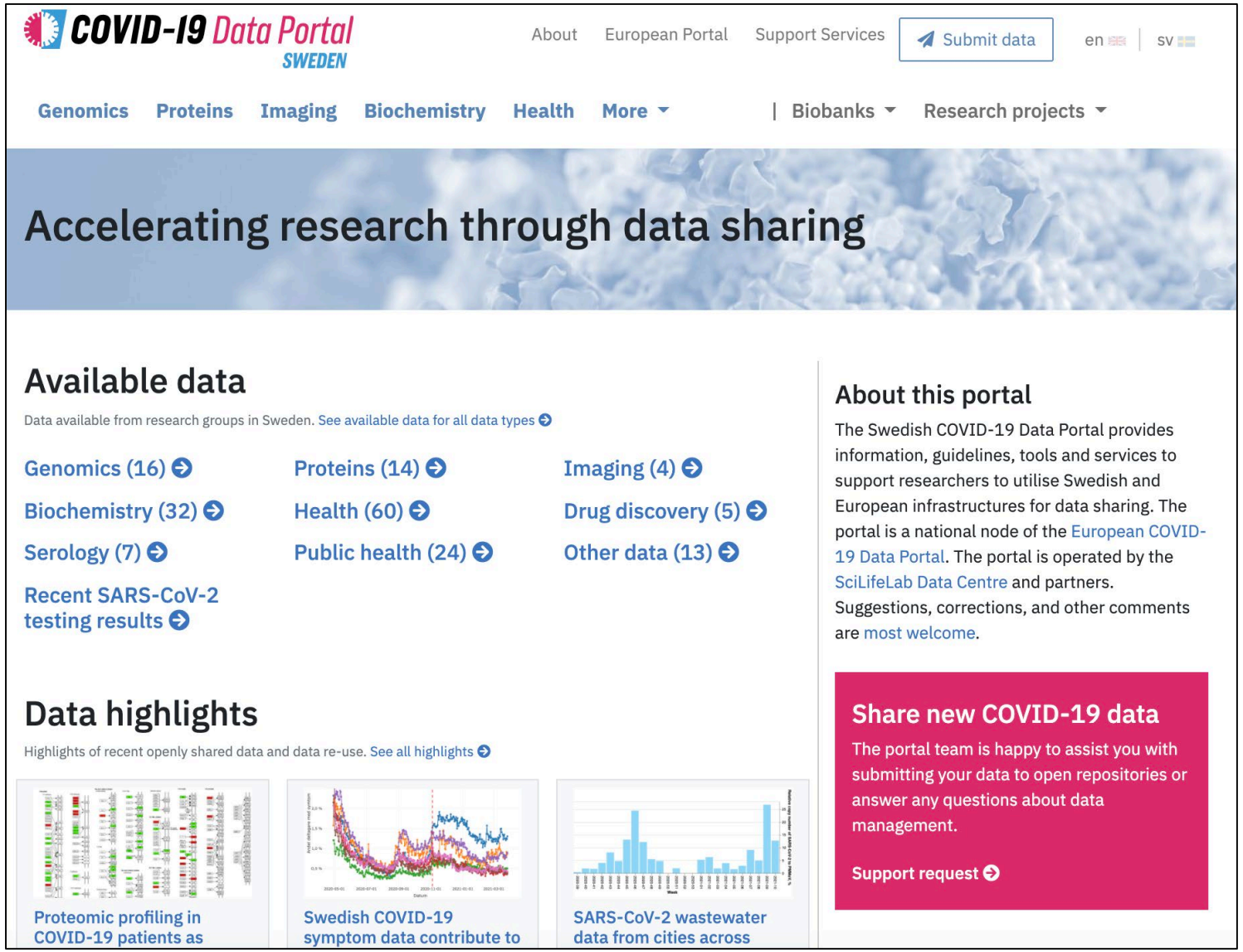
Lisbon, 2021-06-07

Johan Rung

SciLifeLab Data Centre

johan.rung@scilifelab.uu.se

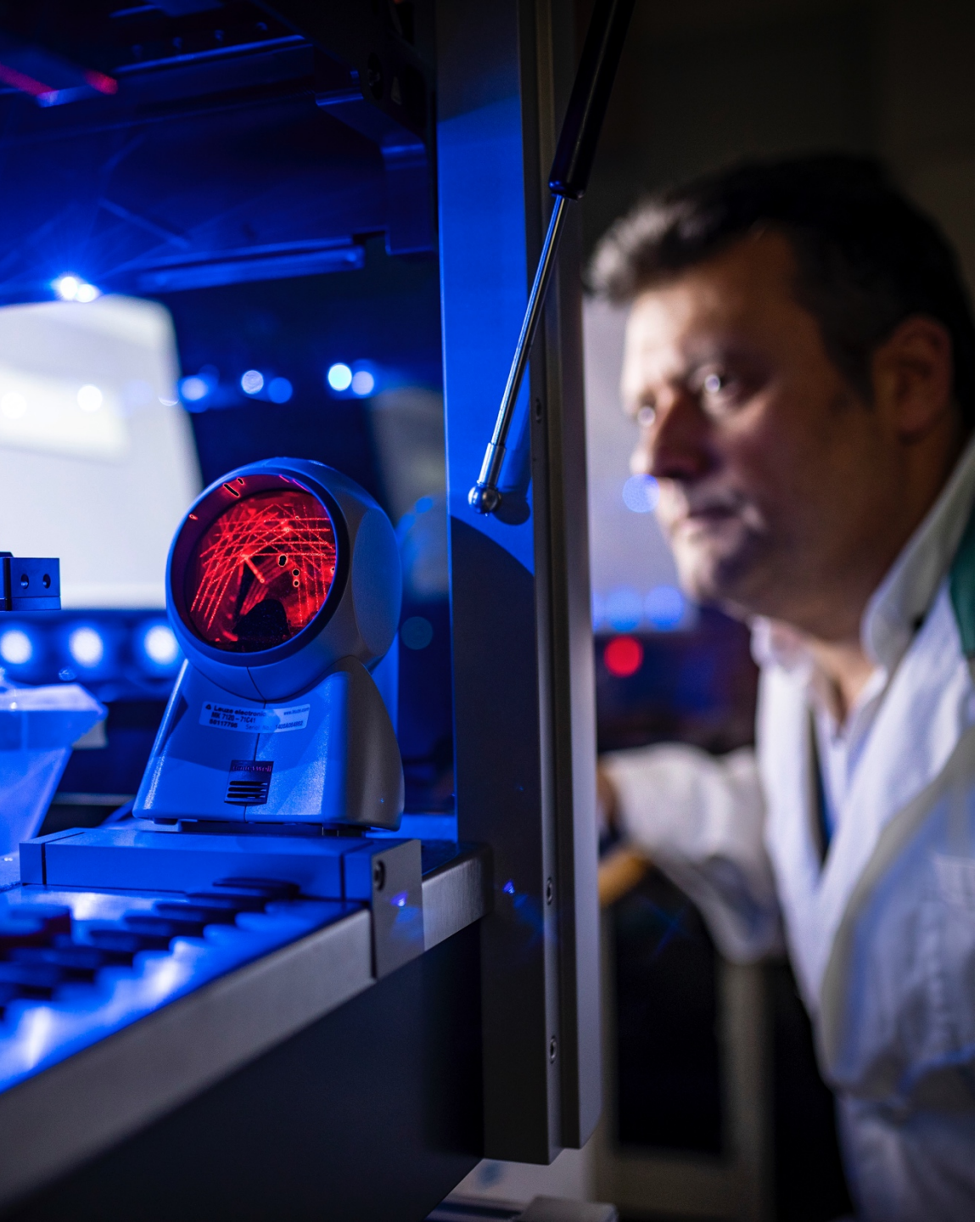
Swedish COVID-19 Data Portal



- The Swedish COVID-19 Data Portal provides information, guidelines, tools and services to support researchers utilizing Swedish and European infrastructures for data sharing, in particular the European COVID-19 Data Portal
- Assignment from Swedish Research Council to SciLifeLab Data Centre
- Data management support for Swedish COVID-19 research programs
- Focus on FAIR data sharing
- Launched June 2020

SciLifeLab

https://covid19dataportal.se



What is SciLifeLab?

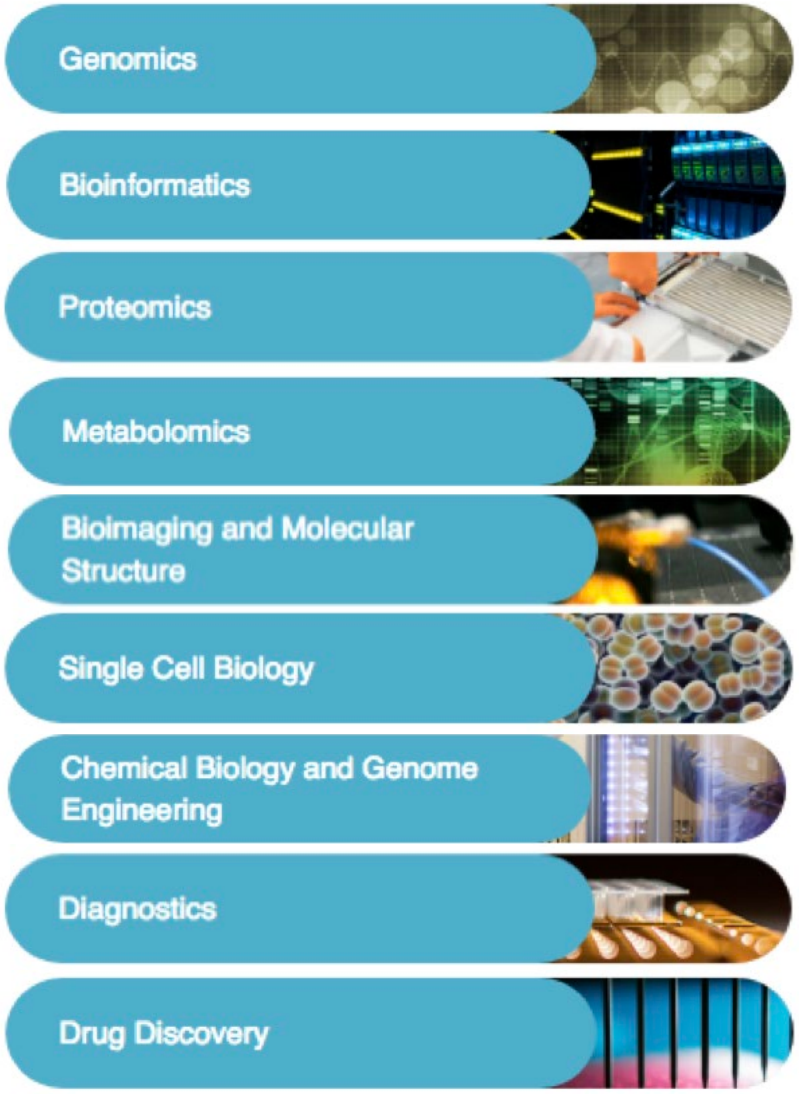
Government appointed mission as a **national research infrastructure in life sciences: 2014-**

Coordinator of a new national 12-year 3 BSEK KAW program of **data-driven life science (DDLS) 2021-**

Research collaboration in life sciences (SFO) across the four host universities (KTH, SU, KI and UU) in Mälardalen **(2010-)**

A hub for the advancement of molecular biosciences in Sweden

Science for Life Laboratory: National infrastructure with 10 platforms, 40 facilities



Bioinformatics

- Support, Infrastructure and Training (G, Li, Lu, S, U, Um)
- Compute and Storage (U)
- BioImage Informatics (U, S)
- AIDA Data Hub (Li)

Genomics

- National Genomics Infrastructure (S, U)
- Ancient DNA (U)
- Microbial Single Cell Genomics (U)

Clinical Genomics

- Clinical Genomics Göteborg
- Clinical Genomics Linköping
- Clinical Genomics Lund
- Clinical Genomics Stockholm
- Clinical Genomics Umeå
- Clinical Genomics Uppsala
- Clinical Genomics Örebro

Metabolomics

- Swedish Metabolomics Centre (Um)
- Exposomics (S)

Clinical Proteomics and Immunology

- Autoimmunity and Serology Profiling (S)
- Translational Plasma Profiling (S)
- Proximity Proteomics (U)
- Mass Cytometry (Li, S)
- Proteogenomics (S)
- Glycoproteomics (G)

Single Cell and Spatial Biology

- Eukaryotic Single Cell Genomics (S)
- Spatial Proteomics(S)
- In Situ Sequencing (S)
- National Resource for Mass Spectrometry Imaging (U)
- Advanced FISH Technologies (S)
- Spatial Transcriptomics (S)

Cellular and Molecular Imaging

- Advanced Light Microscopy (S, G, Um)
- Cryo-EM (S, Um, G, Lu, U)

Integrated Structural Biology

- Swedish NMR Centre (G)
- Structural Proteomics (Lu)

Functional Biology and Target Discovery

- Chemical Biology Consortium Sweden (S, Um)
- Chemical Proteomics (S)
- Center for CRISPR-based Functional Genomics (S)
- Genome Engineering Zebrafish (U)

Drug Discovery and Development

- ADME (Absorption, Distribution, Metabolism, Excretion) of Therapeutics (UDOPP) (U)
- Biochemical and Cellular Assay (S)
- Biophysical Screening and Characterization (U)
- Human Antibody Therapeutics (Lu, S)
- In Vitro and Systems Pharmacology (U)
- Medicinal Chemistry – Hit2Lead (S)
- Medicinal Chemistry – Lead Identification (U)
- Protein Expression and Characterization (S)

SciLifeLab and Wallenberg National Program for Data-Driven Life Science

12 years, 3.1 BSEK, 11 partners

Knut och Alice Wallenbergs Stiftelse

SciLifeLab

WALLENBERG CENTRES FOR MOLECULAR MEDICINE

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

WACQT | Wallenberg Centre for Quantum Technology

CHALMERS

GÖTEBORGS UNIVERSITET

Karolinska Institutet

KTH VETENSKAP OCH KONST

LINKÖPINGS UNIVERSITET

LUNDS UNIVERSITET

SLU

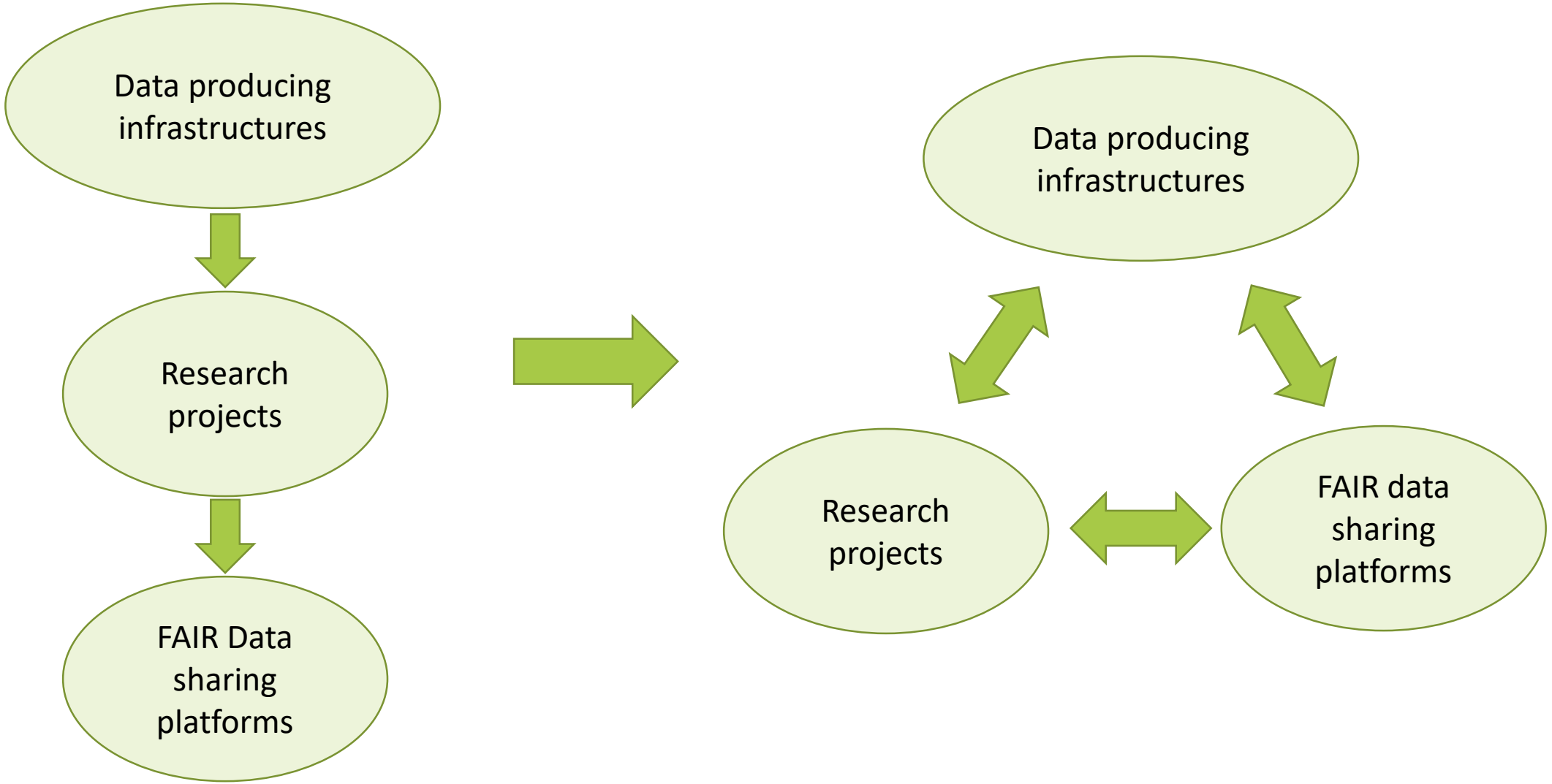
Stockholms universitet

UMEÅ UNIVERSITET

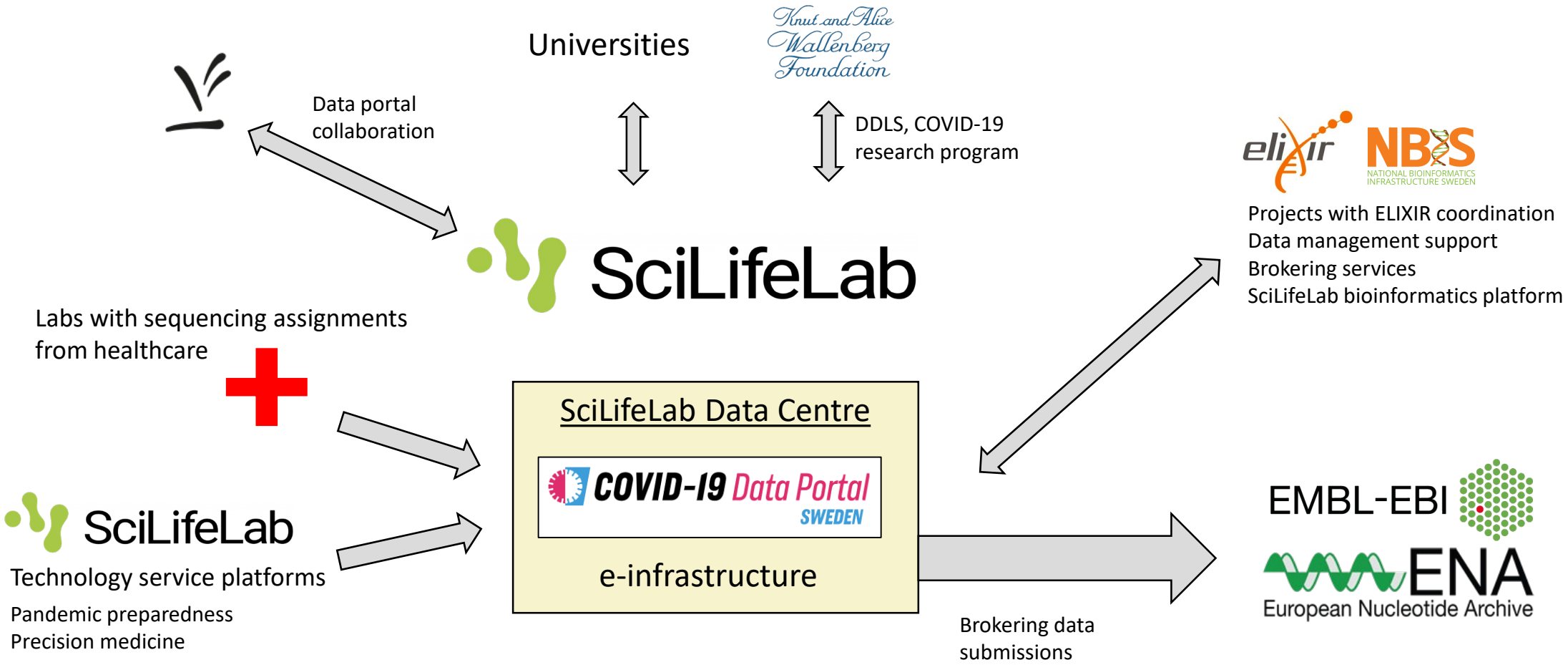
Naturhistoriska riksmuseet

UPPSALA UNIVERSITET

Active role of data sharing in data-driven life science



Swedish COVID-19 data portal

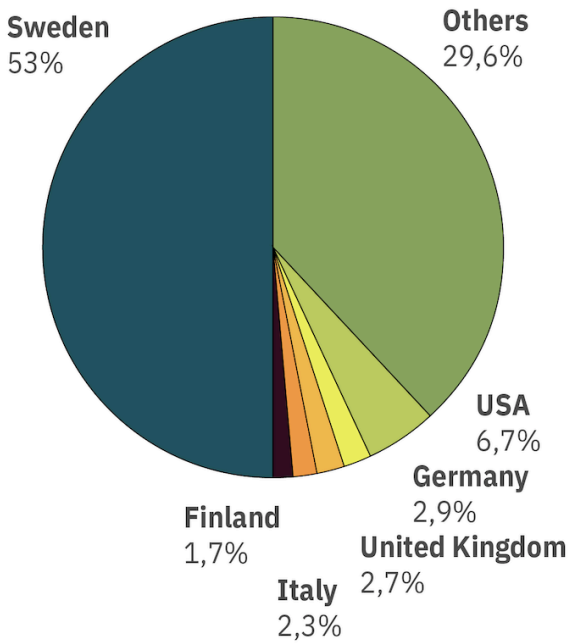


Interoperability between healthcare regions complicated in Sweden.
Meta-data collection varies between regions and data flow from healthcare to research not straightforward.
Obstacles also include consent handling, legal and regulatory obstacles, e.g. longitudinal studies integrating health data and research data.

Swedish COVID-19 Data portal

20.000+
visitors to the portal

100.000+
pageviews of the portal



100+
publications openly sharing data tracked

20+
news items with focus on shared data published

1300+
articles and preprints from Swedish researchers tracked.

200+
research projects funded by major funders tracked

25+
sample collections in biobanks across Sweden in the database

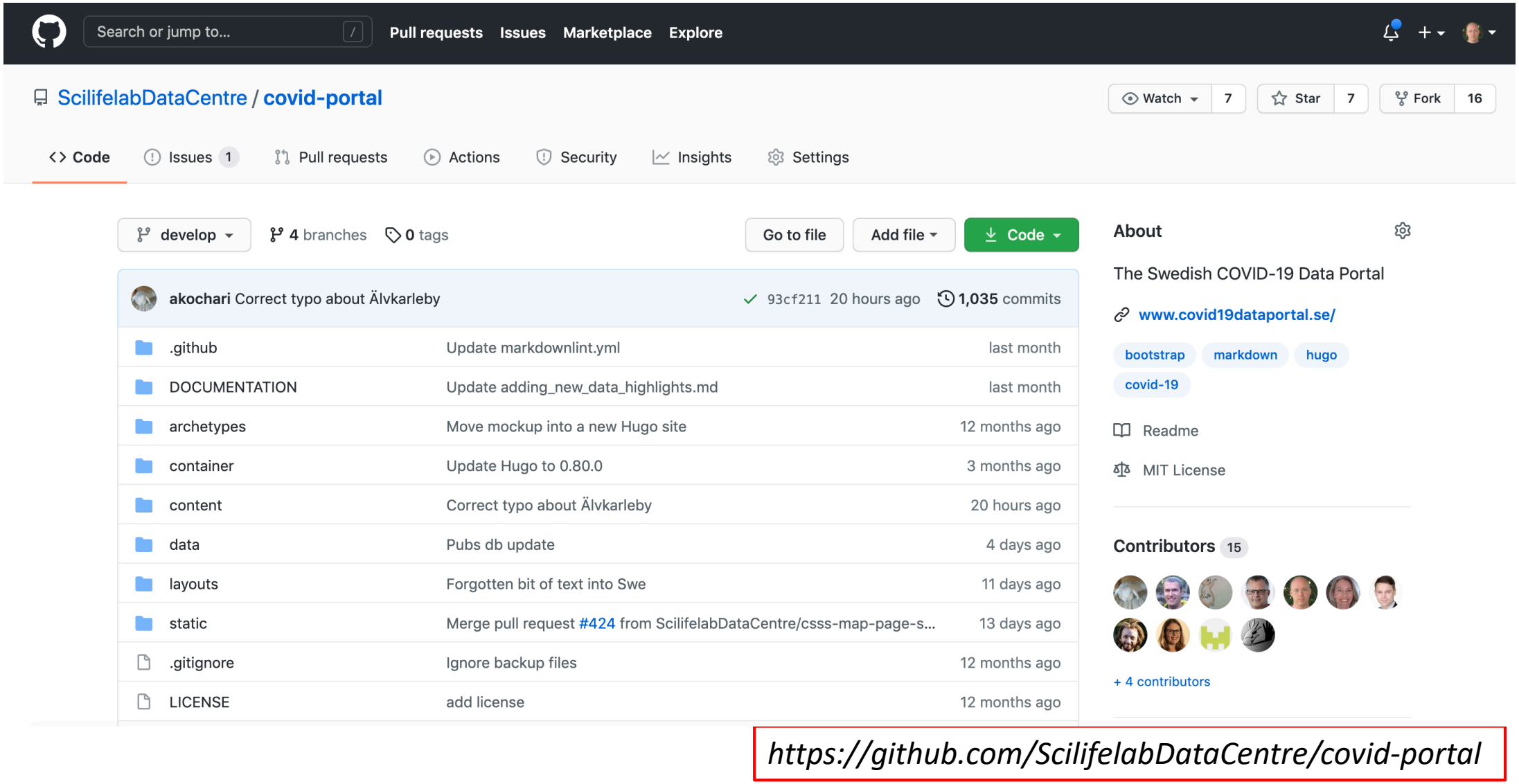
200+
projects in direct contact with portal support team regarding data sharing

Next phase of portal:

- Vaccine research
- Pandemic preparedness
- Healthcare
- e-infrastructure

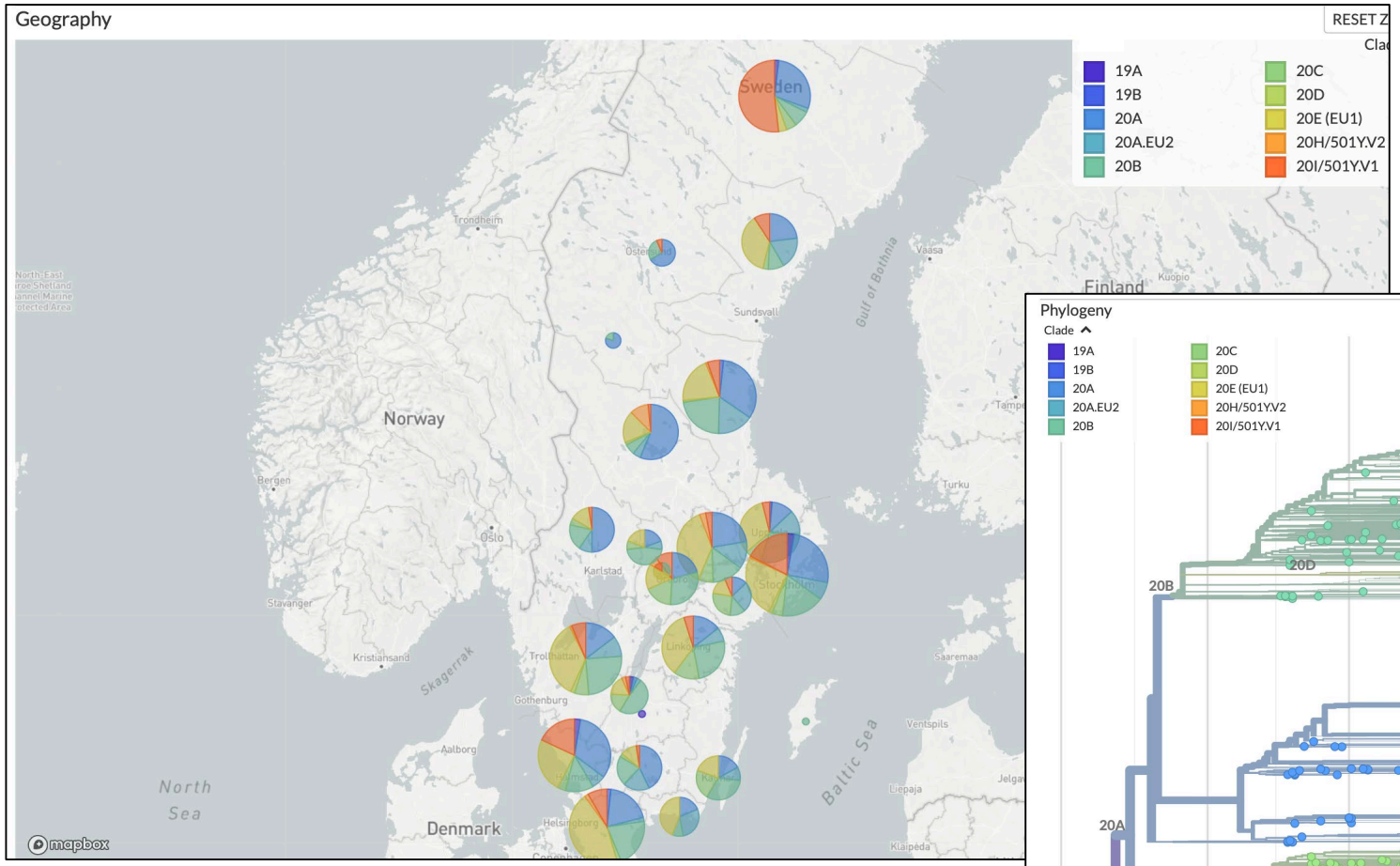
(updated 2021-05-24)

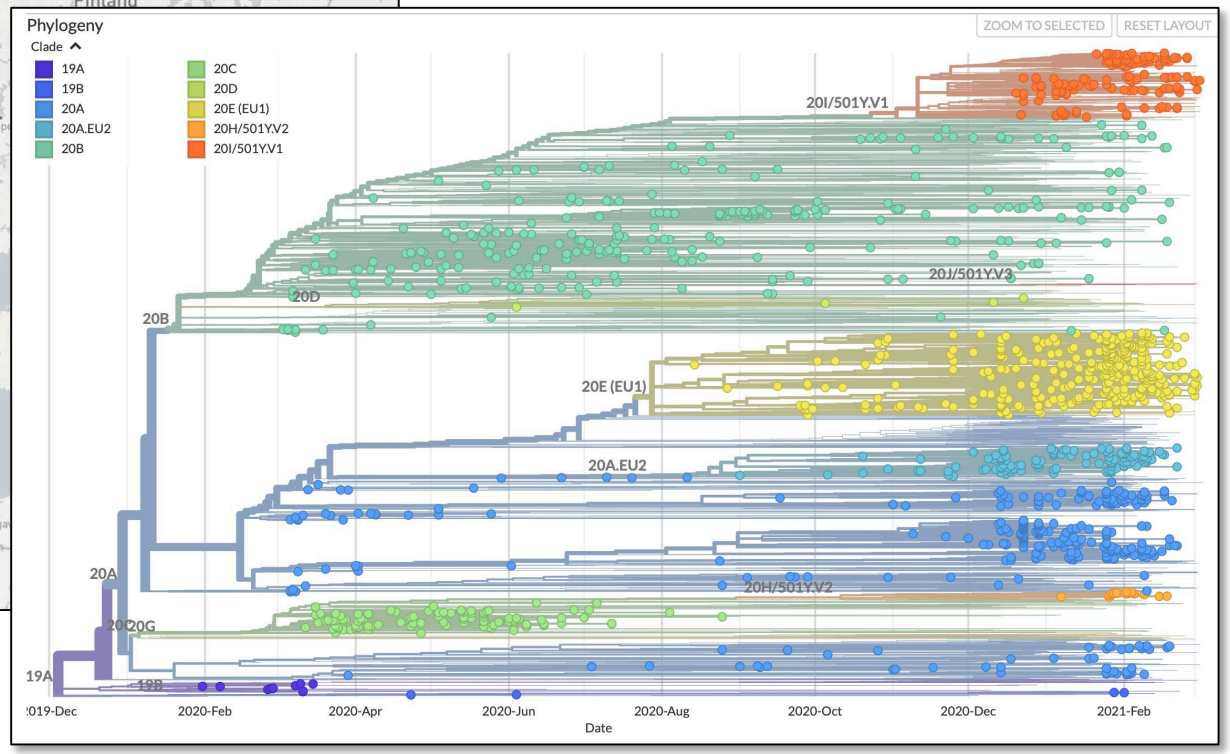
COVID-19 Data Portal source code openly available



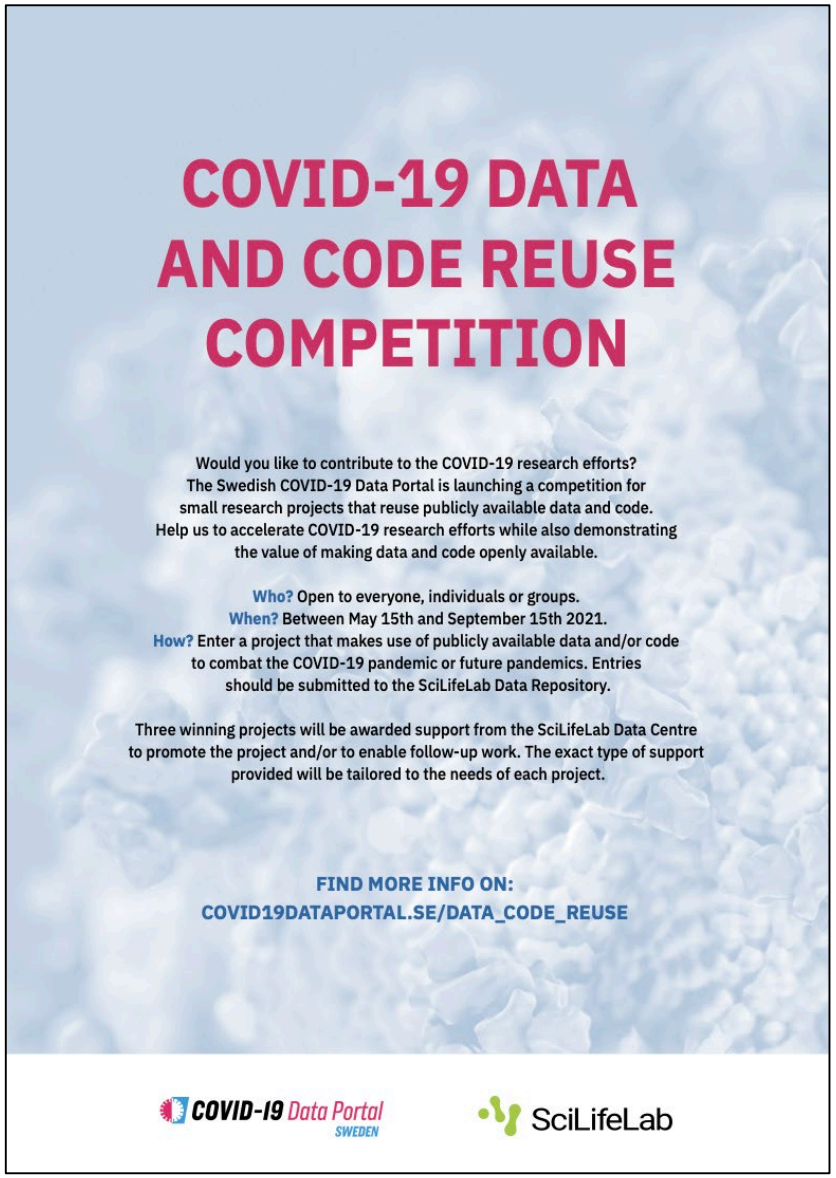
https://github.com/ScilifelabDataCentre/covid-portal

Regional sequencing data





Data and code re-use competition



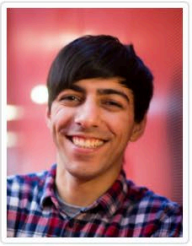
- Demonstrate value in reuseable data and the reach of the COVID-19 Data Portal with students as particular target group
- Open to everyone, even internationally
- Open until Sept 15
- Winning projects will receive support to promote project or enable follow-up work
- Entries to be submitted as SciLifeLab Data Repository items

https://covid19dataportal.se/data_code_reuse/

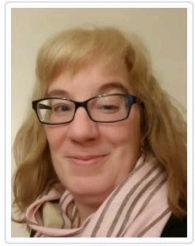
What we learned: critical components

- Rapid international response - EC COVID-19 data platform
- Clear national mandate to develop Data Portal
- Active project support for rapid data sharing and community building
- Existing national and European infrastructure with established collaboration network, existing data infrastructure, and capacity to re-prioritize resources
- Well-organised research programs *requiring* rapid FAIR data sharing
- Direct involvement of researchers in development of interoperable services
- Funding

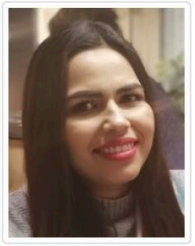
Credits



Arnold Kochari
Webmaster for the COVID-19 Data Portal, SciLifeLab

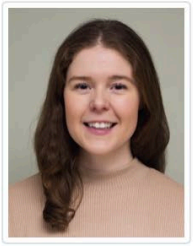


Katarina Öjefors Stark
Coordinator for COVID-19 projects, SciLifeLab



Parul Tewatia
Data steward, SciLifeLab

Liane Hughes
Data Engineer, SciLifeLab



Anna Asklöf
Data Steward, SciLifeLab

Hanna Kultima
Data manager, coordinator, SciLifeLab

Johan Rung
Head of the Data Centre, SciLifeLab

Per Kraulis
System development team leader, data engineer, SciLifeLab

Linus Östberg
Systems developer, system administrator, SciLifeLab

Phil Ewels
Bioinformatics development lead, NGI

Thank you to:

Partner projects
Data contributors

Funding:

Swedish Research Council

SciLifeLab

datacentre@scilifelab.se https://covid19dataportal.se

@scilifelab_DC