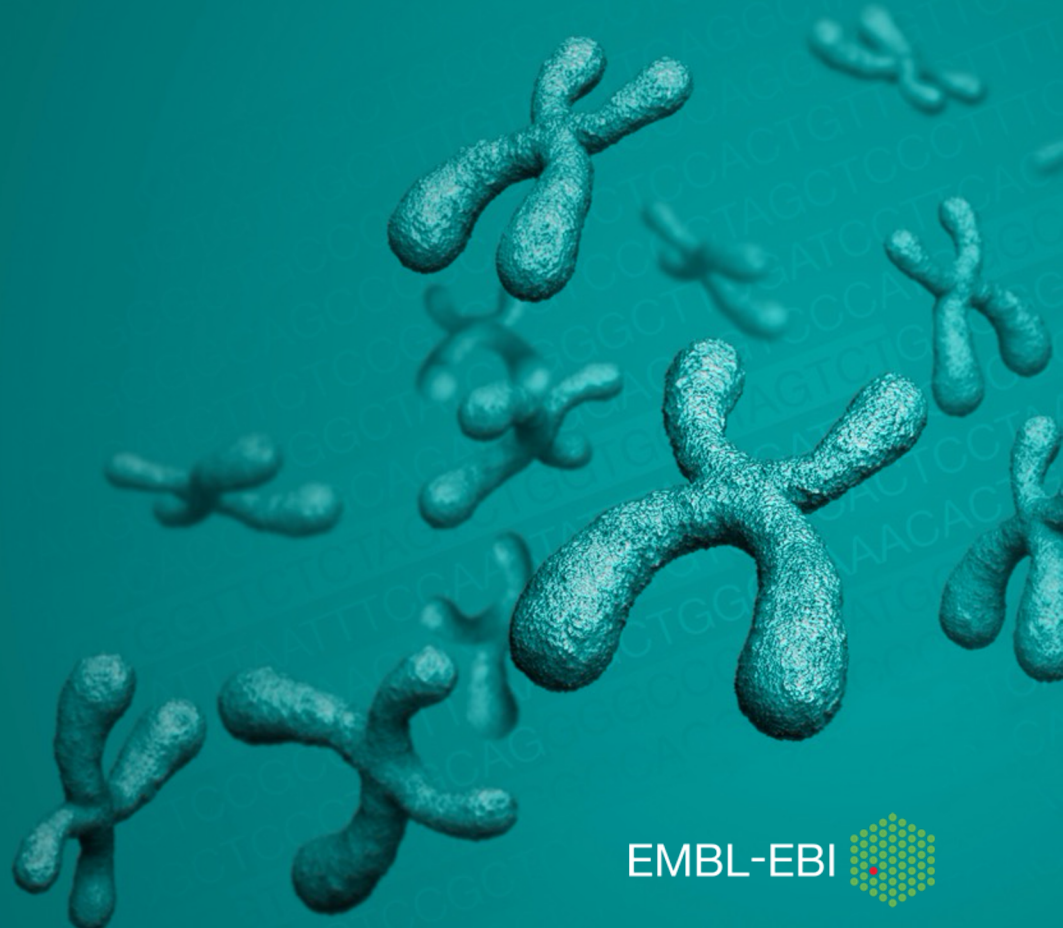
The European COVID-19 Data Platform

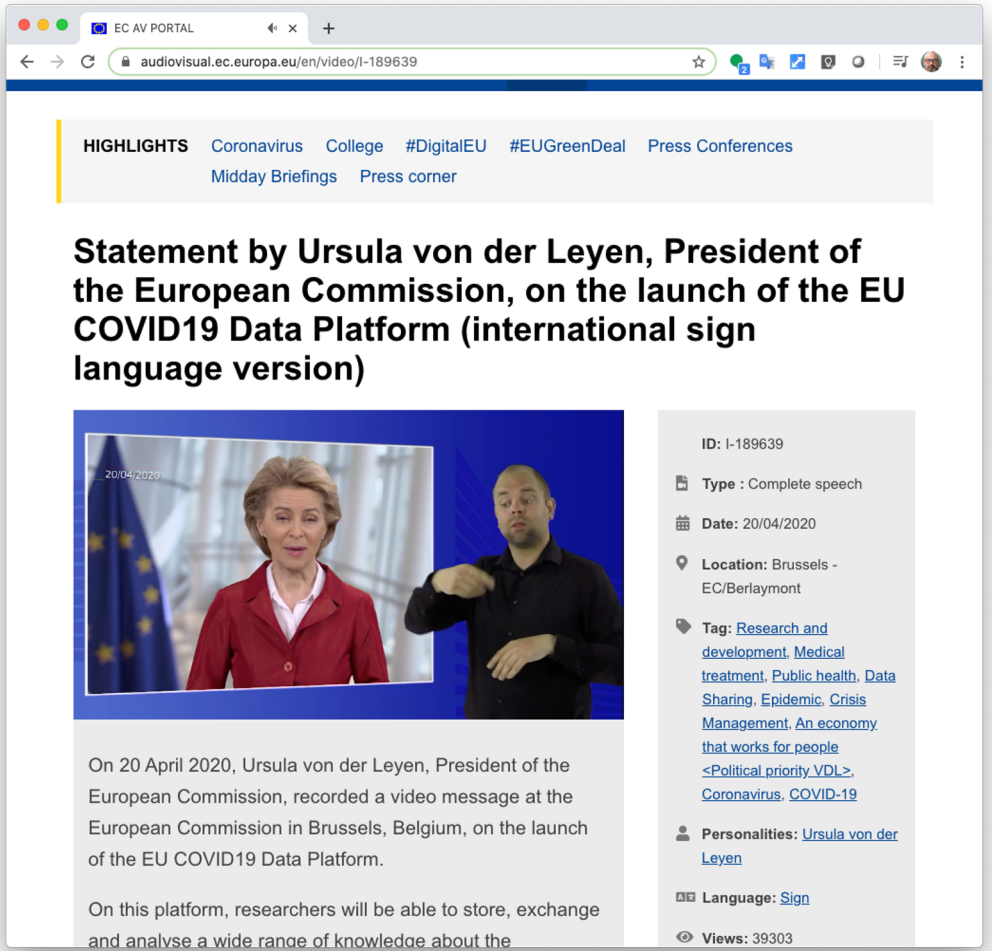
Guy Cochrane



EMBL-EBI

European COVID-19 Data Platform

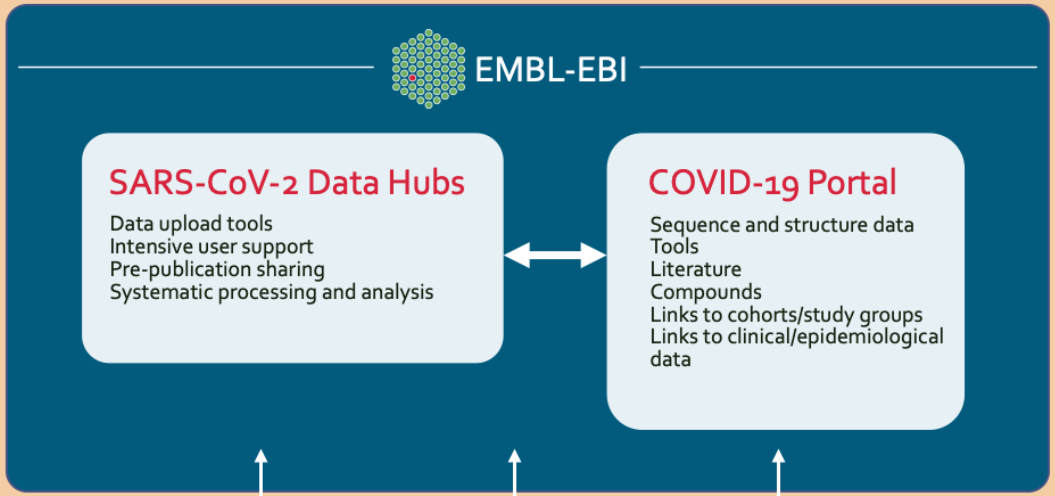
https://www.covid19dataportal.org/



https://audiovisual.ec.europa.eu/en/video/I-189639

EMBL-EBI

European COVID-19 Data Platform

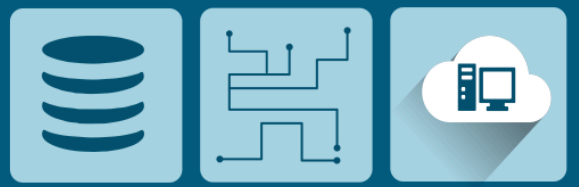


Clinical/epidemiological data where infrastructures do not exist

Sequence Data

Data links

National Infrastructures



erinha — European Research Infrastructure on Highly Pathogenic Agents

instruct ERIC

eatris

EURO BIOIMAGING — Image life, discover the future

EMBRC — European Marine Biological Resource Centre

eu-openscreen — Chemical keys for life's locks

ECRIN — European Clinical Research Infrastructure Network

INFRAFRONTIER — mouse disease models

MIRRI

ISBE

EMPHASIS

ELIXIR CONVERGE

Re CoD ID

EOSC-Life

VEO

CORBEL

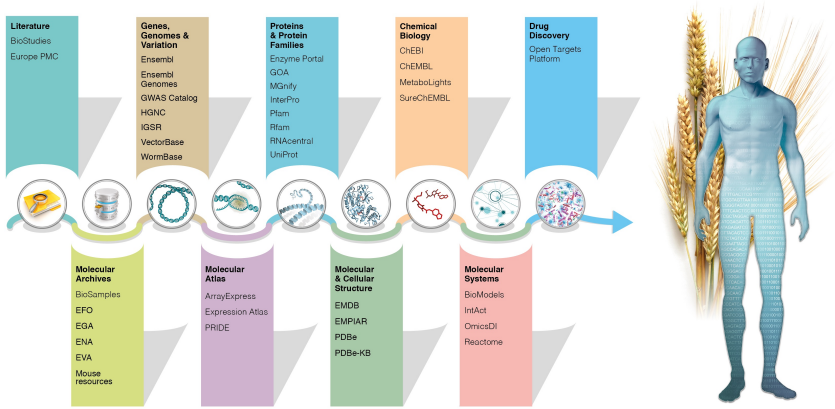
HORIZON 2020 — The Framework Programme for Research and Innovation

EMBL-EBI

ELIXIR

EUROPEAN OPEN SCIENCE CLOUD

Foundations



EMBL-EBI

https://www.ebi.ac.uk/

ELIXIR Members

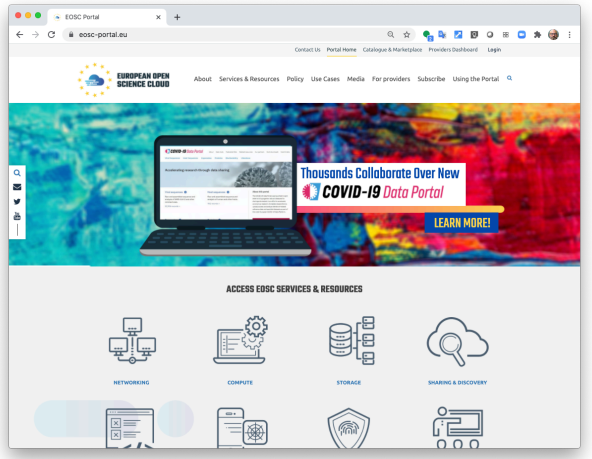
Belgium, Czech Republic, Denmark, EMBL, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Israel, Netherlands, Norway, Slovenia, Portugal, Spain, Sweden, Switzerland, United Kingdom

ELIXIR Observers

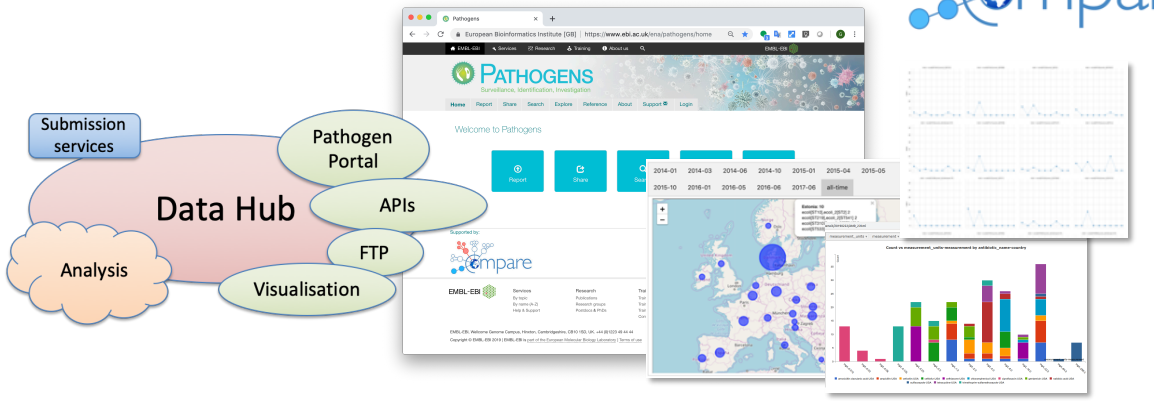
Cyprus

elixir

https://elixir-europe.org/



https://www.eosc-portal.eu/



Amid *et al.* (2019)The COMPARE Data Hubs. *Database : the Journal of Biological Databases and Curation*, 01 Jan 2019, 2019 http://doi.org/10.1093/database/baz136

Erasmus MC

Erasmus Medical Centre, the Netherlands

National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

National Institute for Public Health and the Environment (RIVM), the Netherlands

Eötvös Loránd University

Eötvös Lorand University, Hungary

DTU

Technical University of Denmark (DTU)

UNIVERSITÄTS KLINIKUM HEIDELBERG

Universitätsklinikum Heidelberg, Germany

EUROPEAN OPEN SCIENCE CLOUD

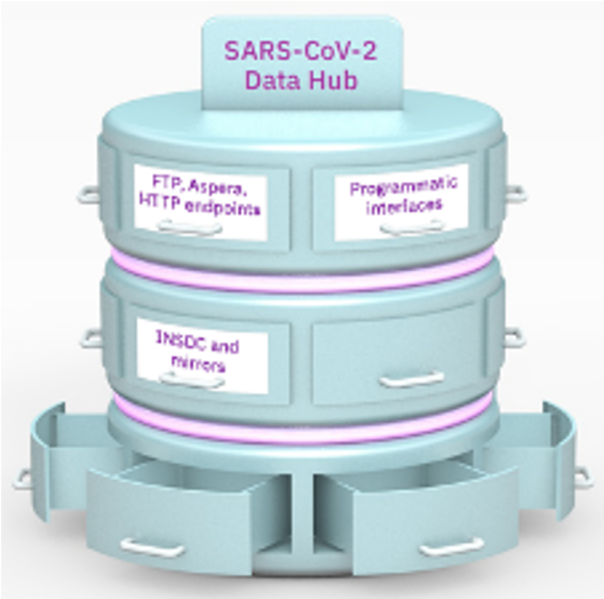
HORIZON 2020
THE FRAMEWORK PROGRAMME FOR RESEARCH AND INNOVATION

EMBL-EBI

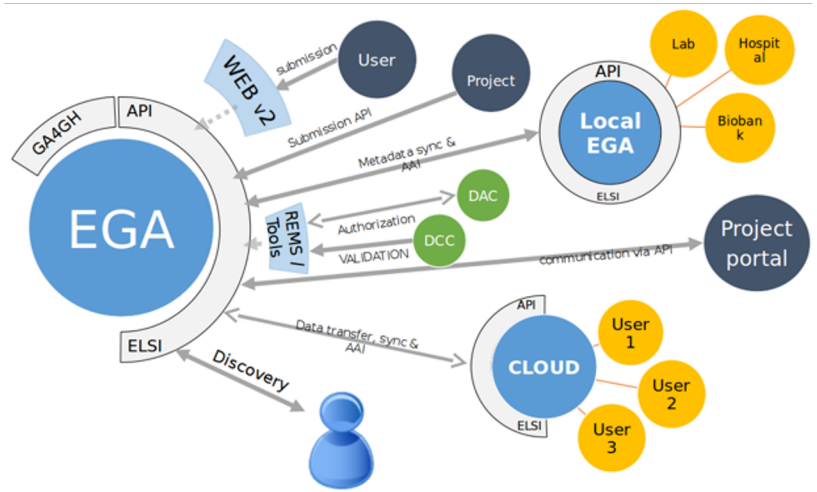
European COVID-19 Data Platform

diverse data – rapid sharing – degrees of completeness – synergy and cross-fertilisation

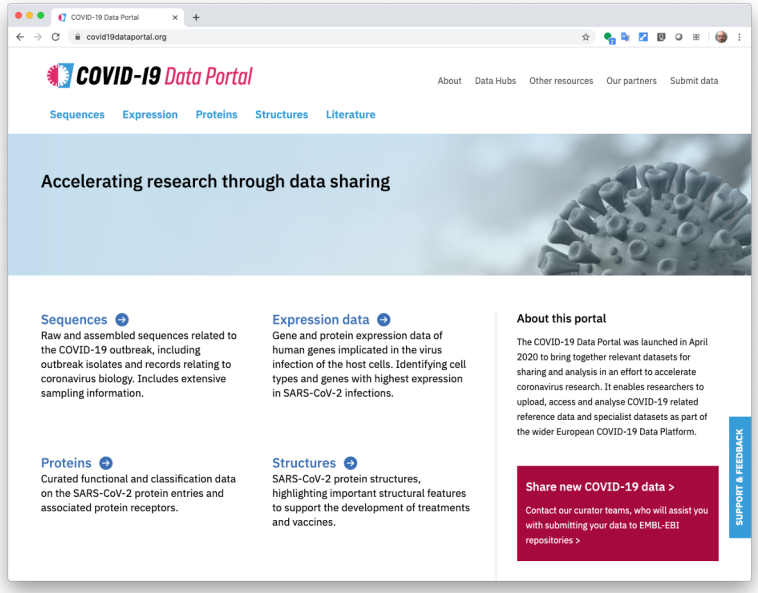
SARS-CoV-2 Data Hubs



Federated European Genome-phenome Archive



COVID-19 Data Portal



https://www.covid19dataportal.org/

EMBL-EBI

Priority I: data integration

EMBL-EBI

Data integration

Viral Sequences

All (152,407)
Sequences (23,861)
Raw reads (65,323)
Sequenced samples (50,342)
Studies (167)
Genes (22)
Browser (1)
Variants (12,691)

Host Sequences

All (973)
Human studies (controlled access) (3)
Human reads (consented for full access) (712)
Other species reads (250)
Association studies (8)

Expression

All (69)
Gene expression (4)
Gene expression experiments (23)
Single cell expression (2)
Single cell expression experiments (14)
Protein expression experiments (26)

Proteins

All (611)
Protein sequences (53)
Protein families (126)
Protein structures - Knowledge Base (7)
Protein structures (260)
Electron microscopy density maps (165)

Biochemistry

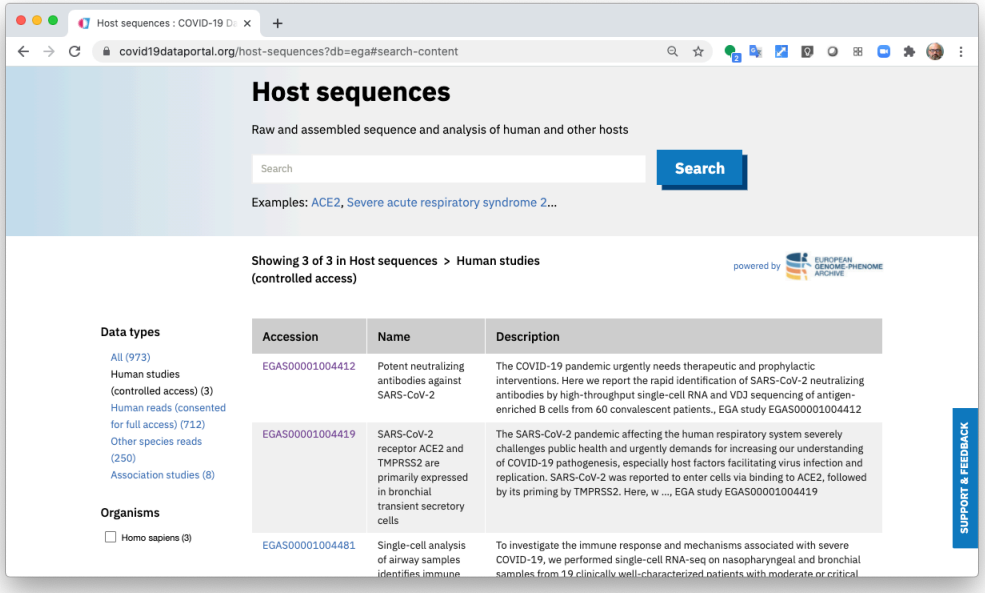
All (1,853)
Pathways (16)
Interactions (1,415)
Complexes (24)
Compound document (8)
Drug targets (390)

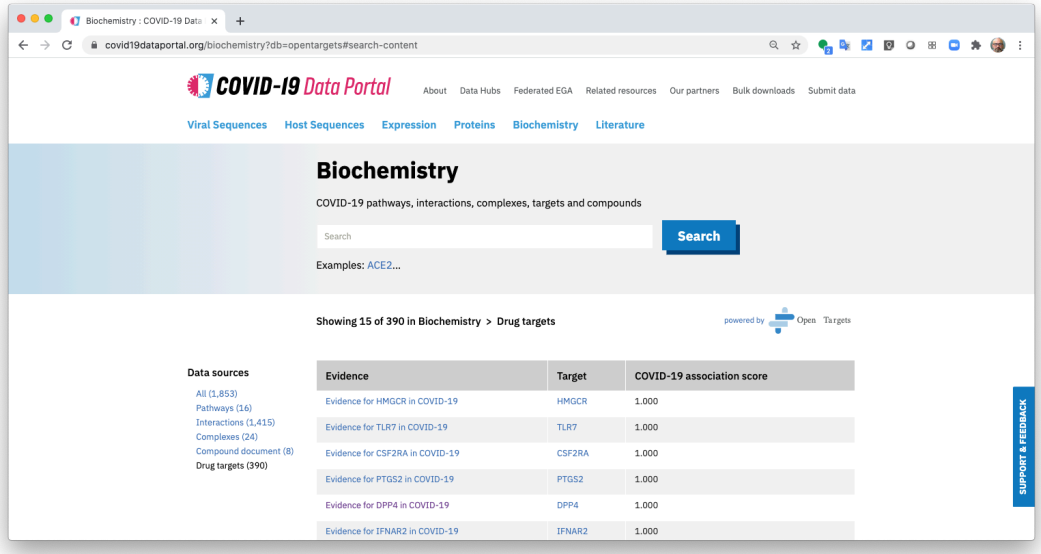
Literature

All (176,427)
Coronaviruses (90,722)
Diseases (79,381)
Related viruses and diseases (2,007)
Genes, receptors and antibodies (4,317)

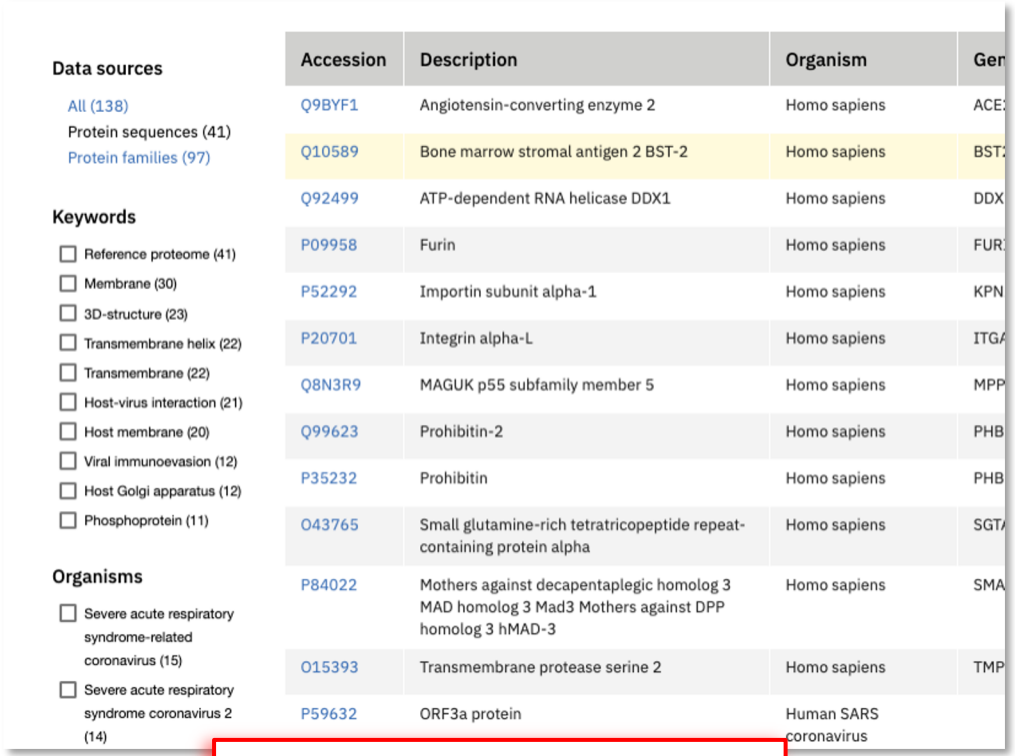
Organisms

- Mus musculus (154)
- Macaca mulatta (56)
- Chlorocebus aethiops aethiops (16)
- Macaca fascicularis (8)
- Homo sapiens (8)
- Chlorocebus sabaeus (6)
- Severe acute respiratory syndrome coronavirus 2 (2)

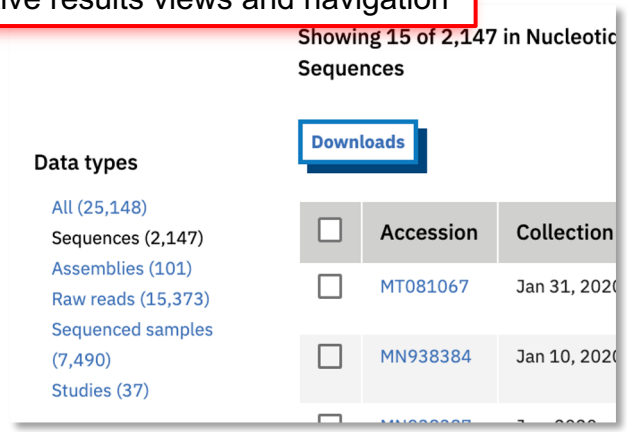




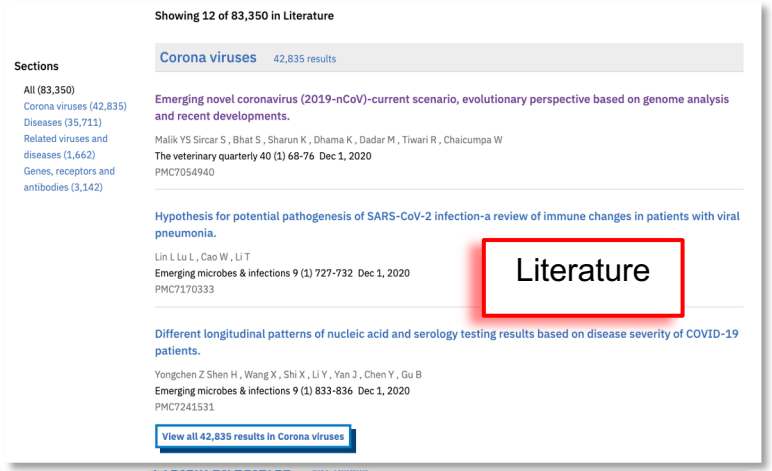
EMBL-EBI



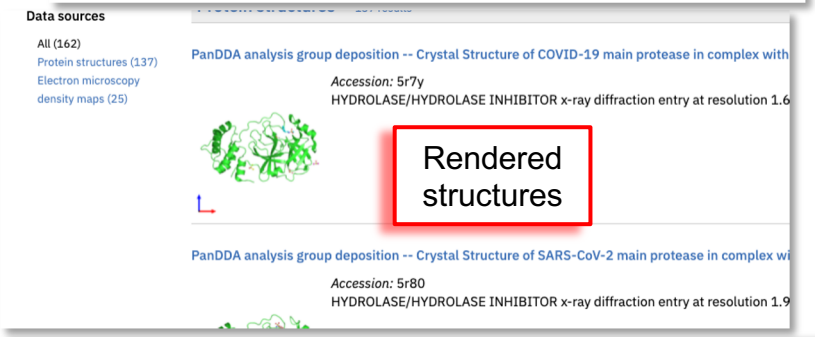
Intuitive results views and navigation



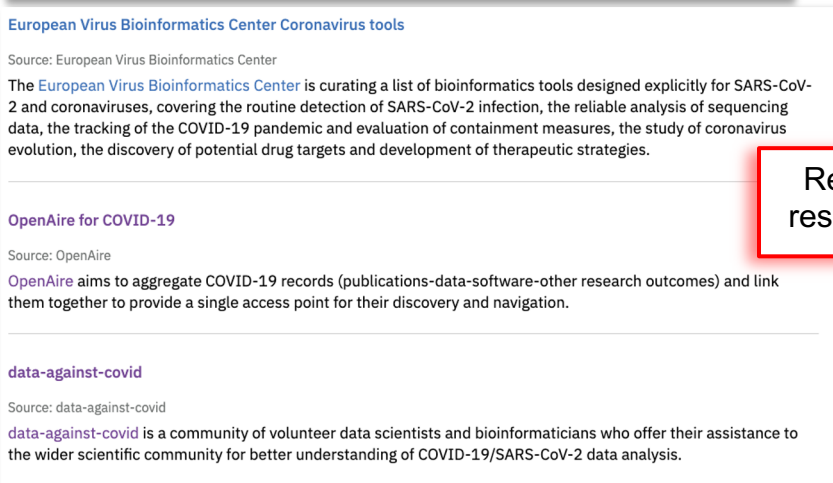
https://www.covid19dataportal.org/



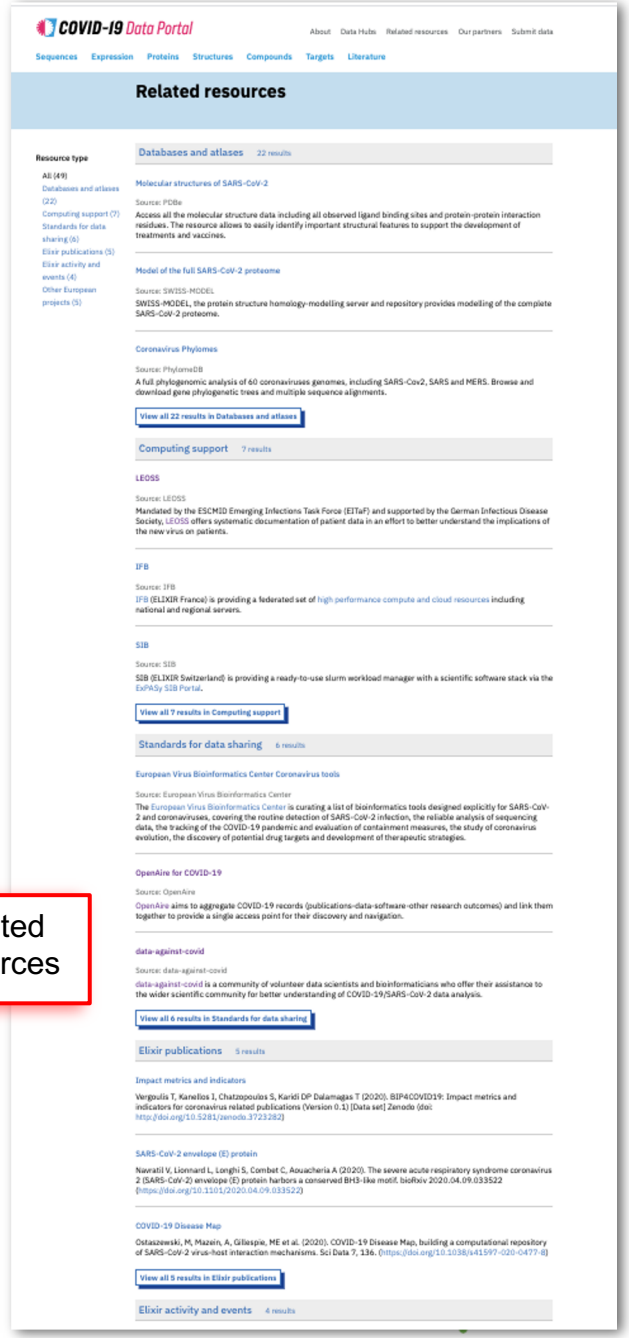
Literature



Rendered structures

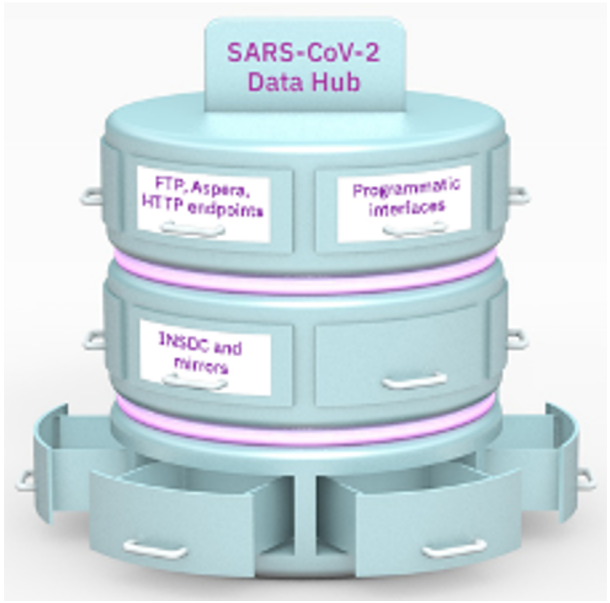


Related resources



Priority II: viral data mobilisation

Viral data mobilization

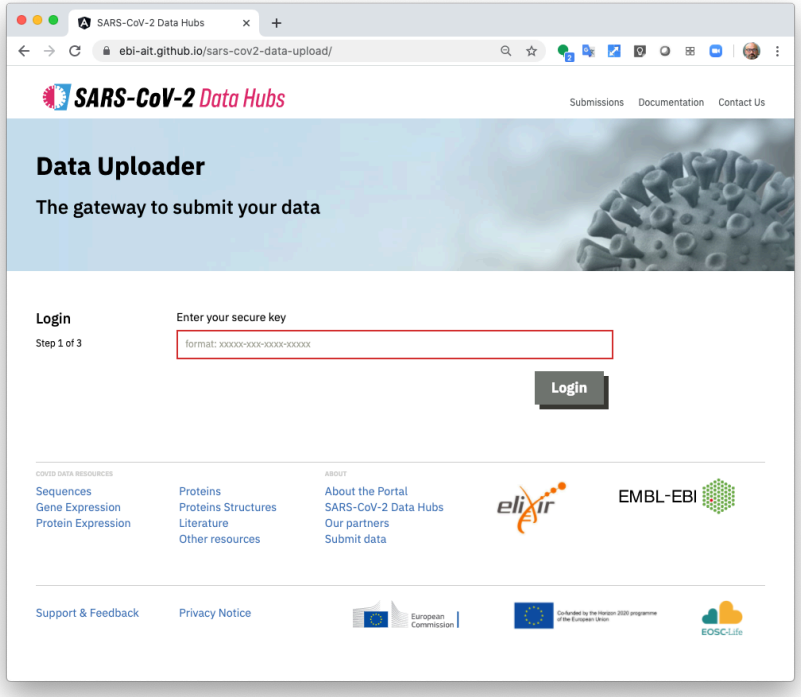


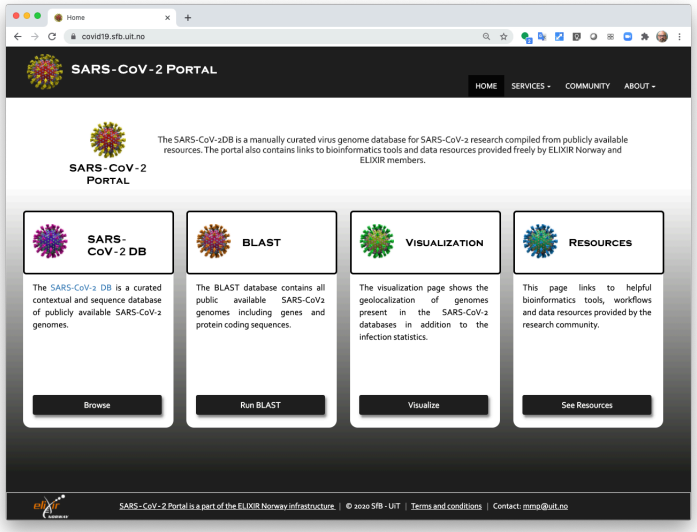
- Raw viral sequence data key to tractable viral variation calling
- Viral variation informs the study of the biology, transmission, spread of the virus
- Required to inform interventions, treatments, vaccine design, etc.

- Ambition to mobilise raw sequence data at scale to support systematic analysis and intercomparability

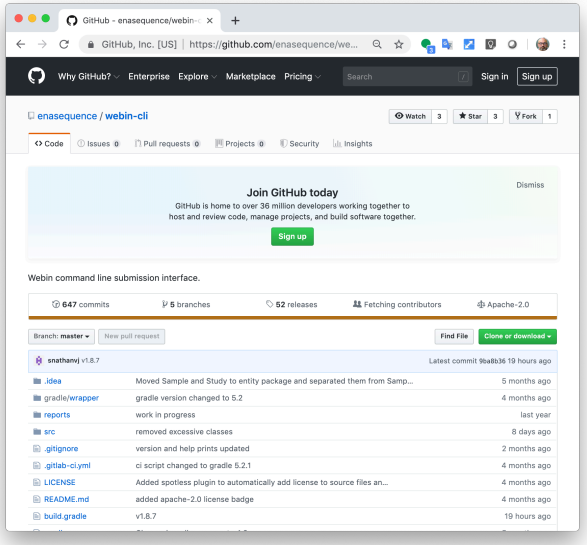
Uploader and user support effort

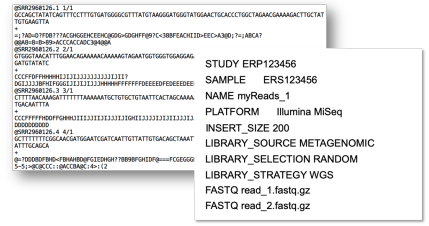
- Tools and services to support data upload
  - SARS-CoV-2 Uploader
  - Existing programmatic and interactive tools
- Extensive support desk
- European campaign
- Standards and compliance support
- Curation, harmonization and update cycles

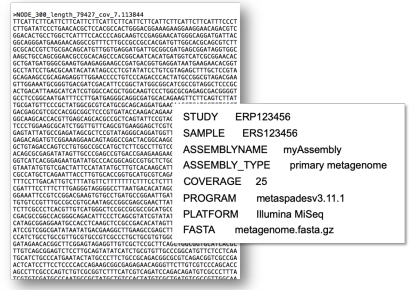




https://covid19.sfb.uit.no/



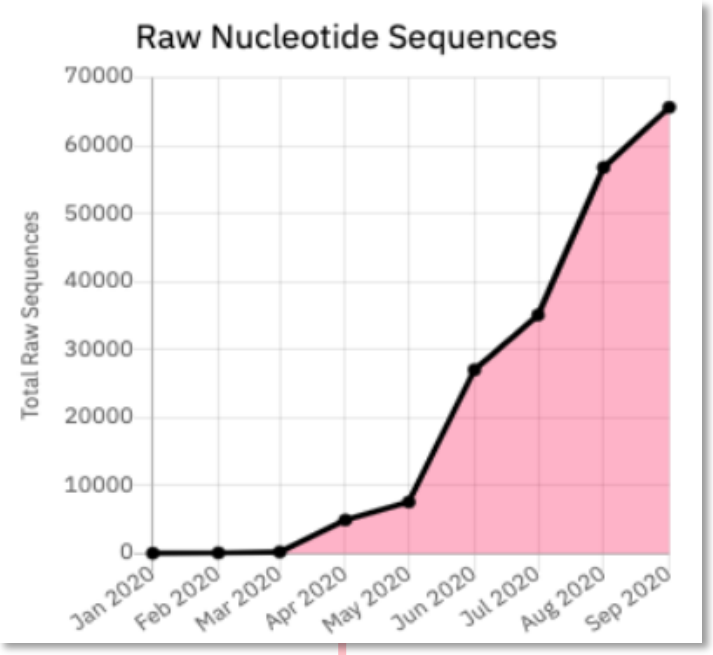


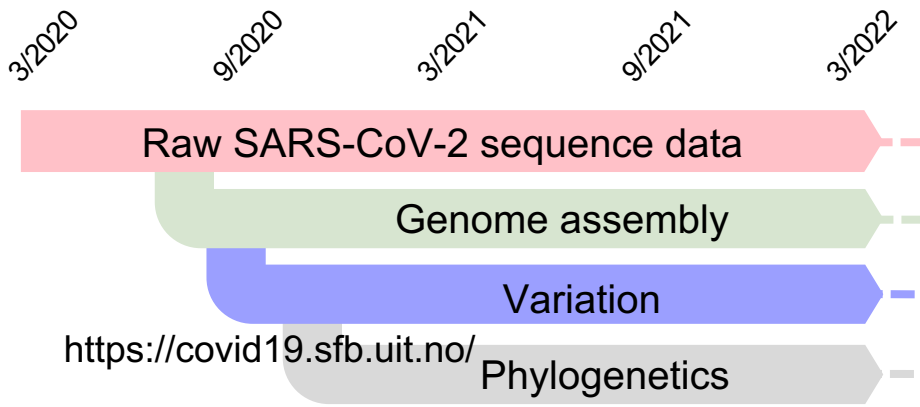


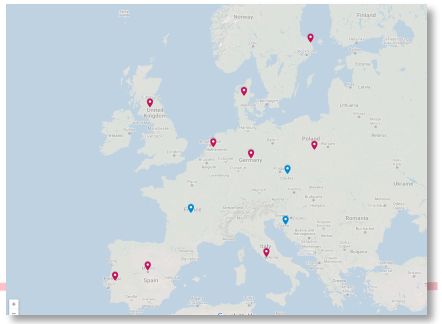
EMBL-EBI

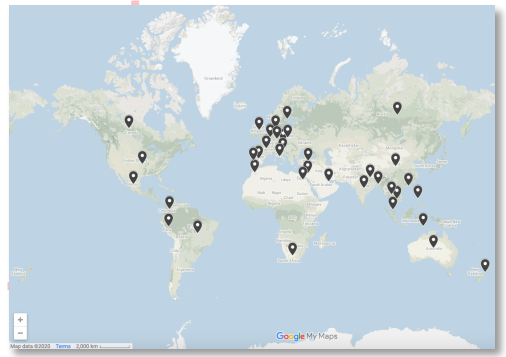
Status at 6-month point

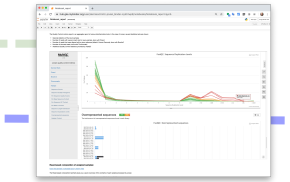
- 65,000 sequenced isolates from 38 countries and 300 institutions
- 16 national SARS-CoV-2 Data Hubs supporting 70% of data
- Curated metadata
  - primer sets
  - sample descriptions with ELIXIR-Norway
- 3 computational workflows (from Erasmus MC and RIVM)
- Assembly product currently in QC, soon to be published
- Promotion to the scientific community

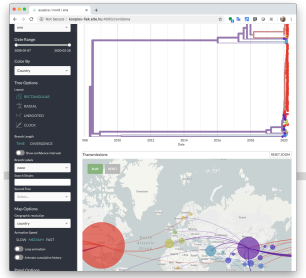






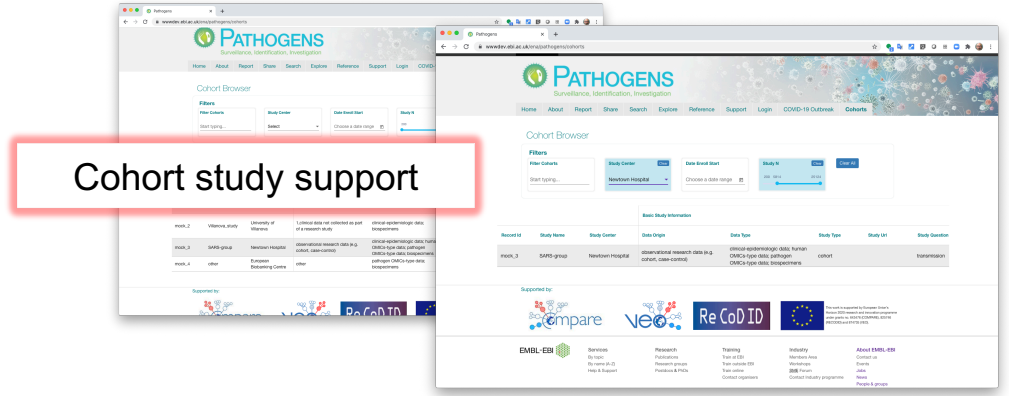




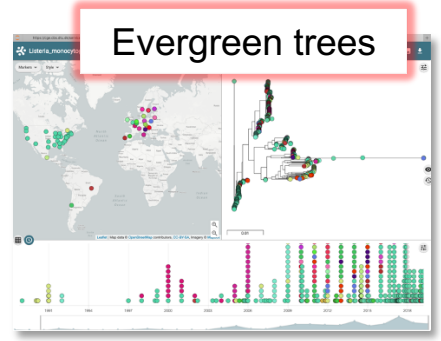


EMBL-EBI

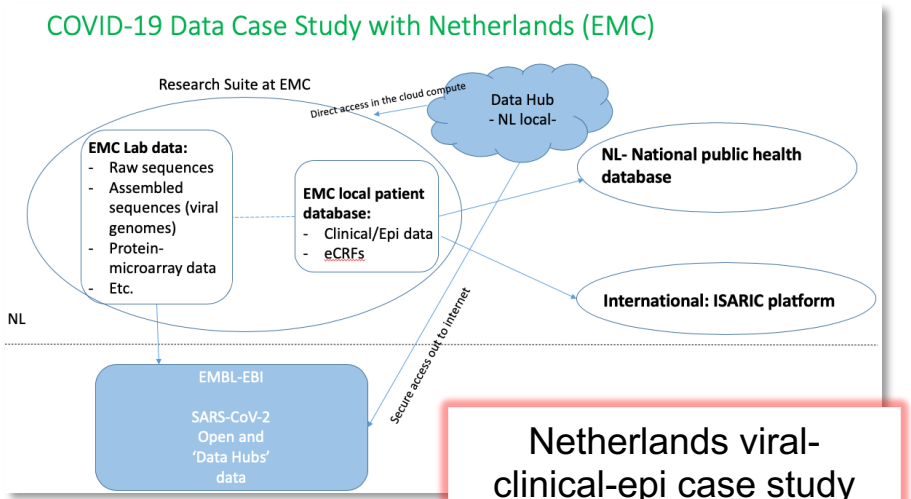
Further work streams



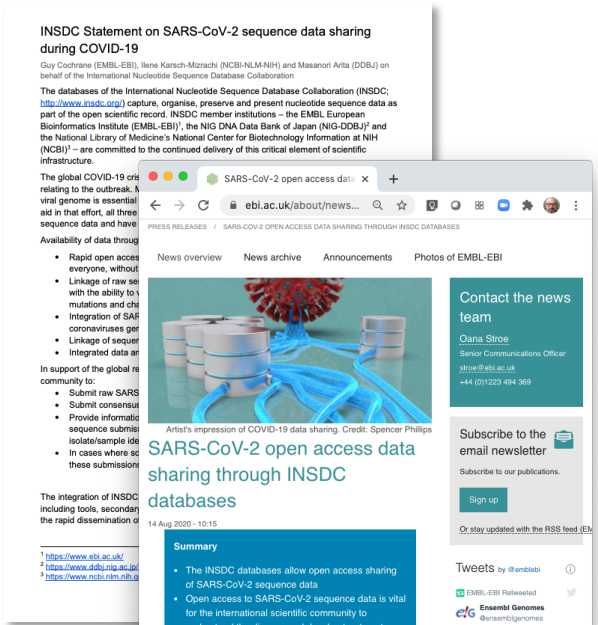
Cohort study support



Evergreen trees

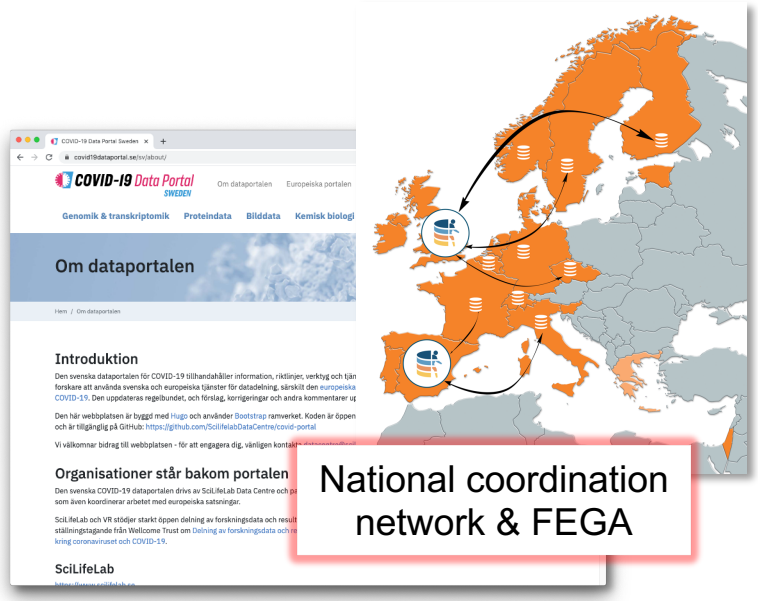


Netherlands viral-clinical-epi case study



Promotion of FAIR practice

https://www.ebi.ac.uk/about/news/announcements/sars-cov-2-open-access-data-sharing-through-insdc-databases



National coordination network & FEGA

https://www.covid19dataportal.se/

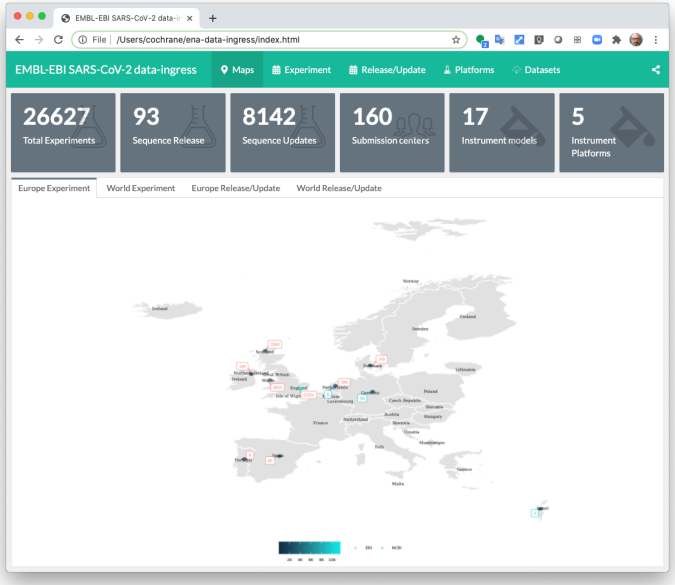
Serology data standards

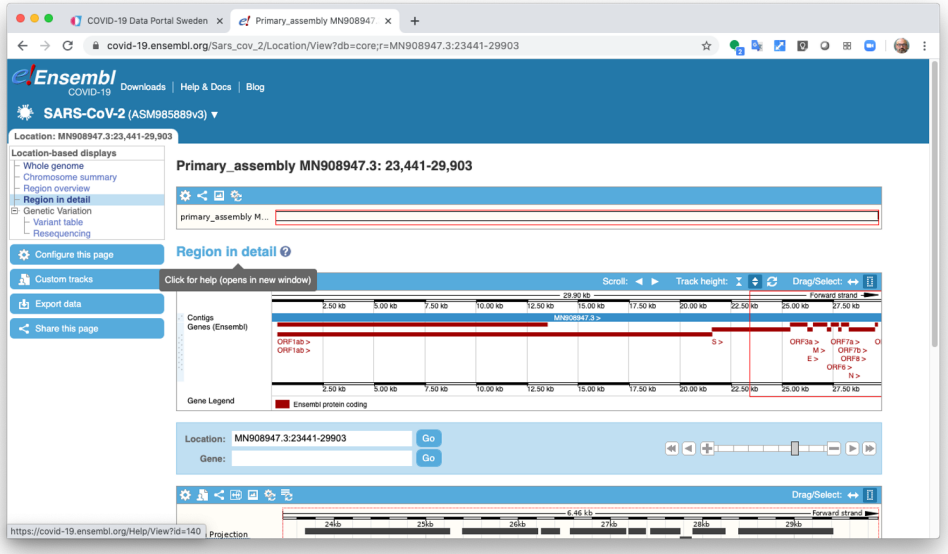
Exploratory work
Data standards
Minimal reporting requirements
Deposition database
Support for
- interoperability across methods
- comparator studies
- engaged communities

EMBL-EBI

Future

- Broader and richer data
- Further tools for data exploration, analysis and visualization
- Phylogenetic integration
- Discovery of cohorts and their data
- Connected viral-human data sets





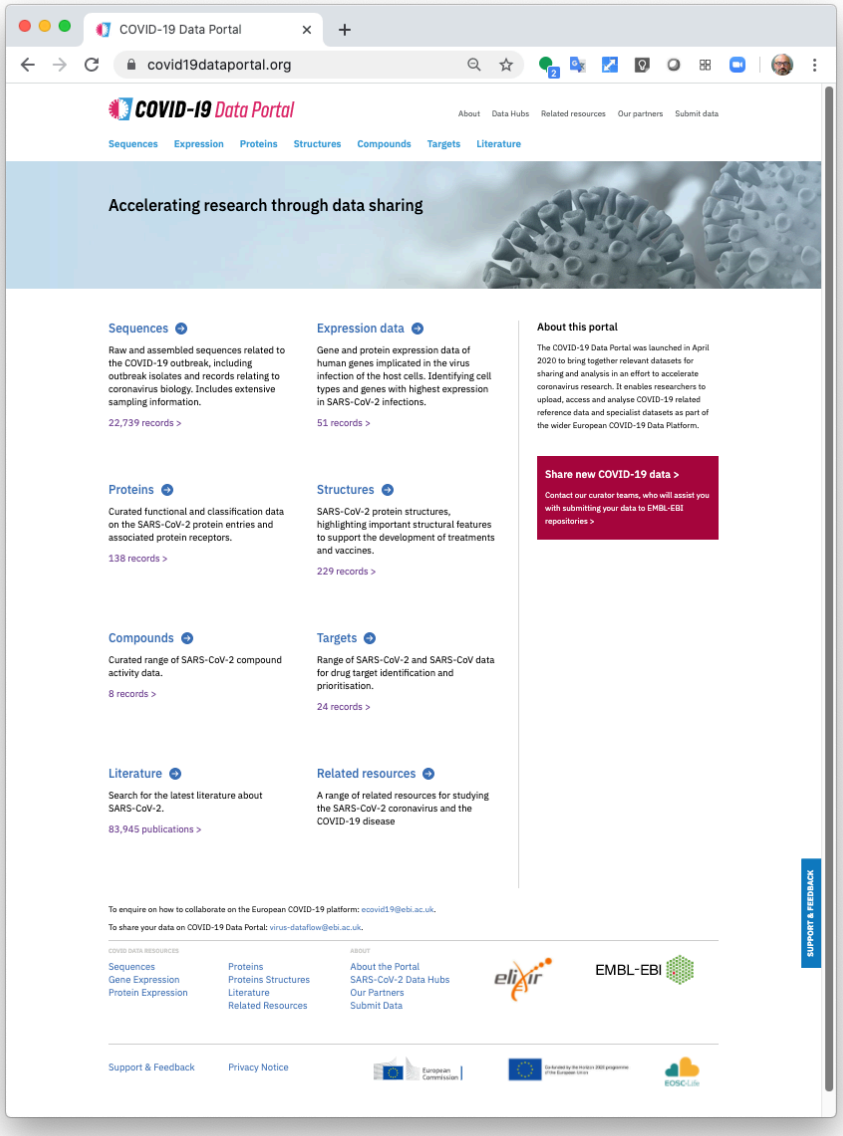
EMBL-EBI

Engagement

- Access data and services via the COVID-19 Data Platform
- Access tools and support for viral data management, analysis and sharing
- Share cohort information and connected data sets

- For national and regional coordinators
  - Explore options for national FEGA activities
  - Join our international stakeholder group

- Contact cochrane@ebi.ac.uk



https://www.covid19dataportal.org/

EMBL-EBI