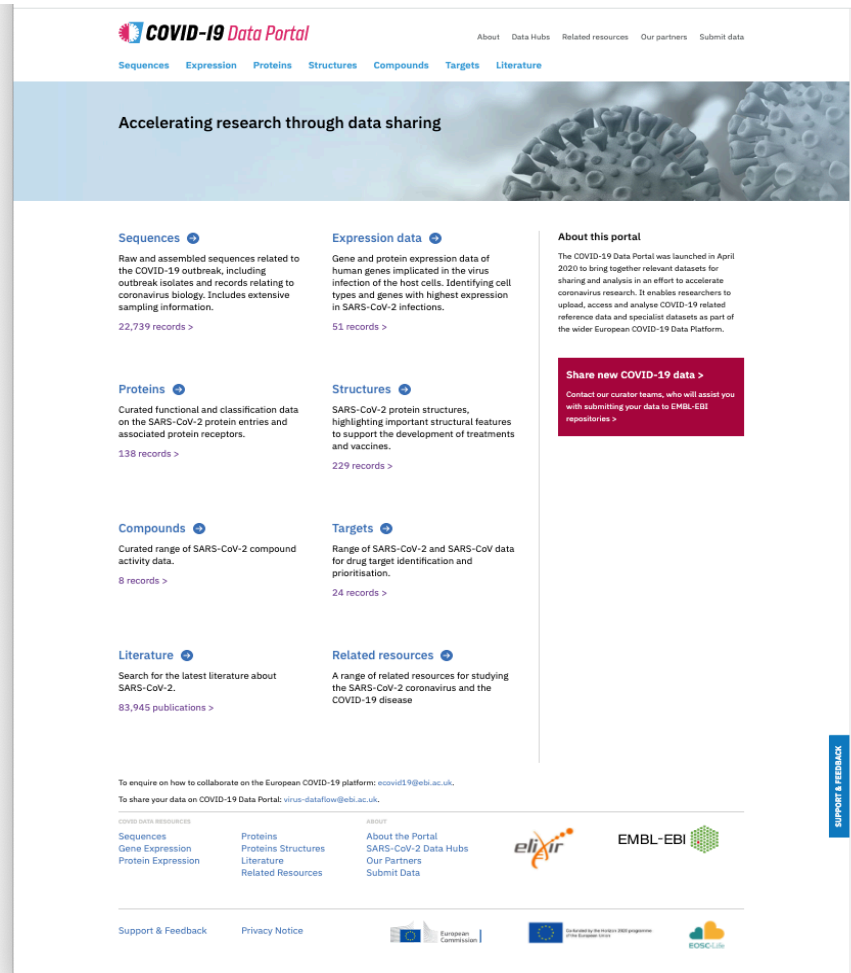# European COVID-19 Data Platform

e-IRG Virtual Workshop under the German EU Presidency

Dec 2nd, 2020

Kostas Repanas

European Commission

Directorate-General for Research and Innovation

Open Science unit

European COVID-19 Data Platform

https://www.covid19dataportal.org/

- Open and rapid access to data, tools and workflows

- Built upon open standards

- Global data coverage and global access

- Enabling diverse research to fight COVID-19

Platform launch

ERAvsCORONA ACTION PLAN: FIRST RESULTS

Acting fast – acting together

European Union

26 June 2020
#Coronavirus

"Three months ago, we were facing a pandemic of unknown proportions. But Europe's scientific talent and innovative capacity gave us hope. Through the ERAvsCorona Action Plan, we jointly mobilised European research and innovation to tackle the crisis, and we did it fast. We joined forces, scaled up our response, and it made us stronger."

Mariya Gabriel, Commissioner for Innovation, Research, Culture, Education and Youth

The ERAvsCorona Action Plan, developed jointly by Commission services and national authorities, features **10 short-term research and innovation actions** to tackle the coronavirus.

In just three months of working together, the Action Plan has delivered results. They include closer coordination of actions taken by the Member States and the Commission, joining forces in providing financial support, creating new funding opportunities, refocusing existing projects, sharing data, setting mechanisms to match great ideas with market opportunities and much more.

It shows that **acting fast and acting together** has big impact.

https://ec.europa.eu/info/sites/info/files/research_and_innovation
/research_by_area/documents/ec_rtd_era-vs-corona-results.pdf

HIGHLIGHTS Coronavirus College #DigitalEU #EUGreenDeal Press Conferences
Midday Briefings Press corner

Statement by Ursula von der Leyen, President of the European Commission, on the launch of the EU COVID19 Data Platform (international sign language version)

ID: I-189639

Type : Complete speech

Date: 20/04/2020

Location: Brussels - EC/Berlaymont

Tag: Research and development, Medical treatment, Public health, Data Sharing, Epidemic, Crisis Management, An economy that works for people <Political priority VDL>, Coronavirus, COVID-19

Personalities: Ursula von der Leyen

Language: Sign

Views: 39303
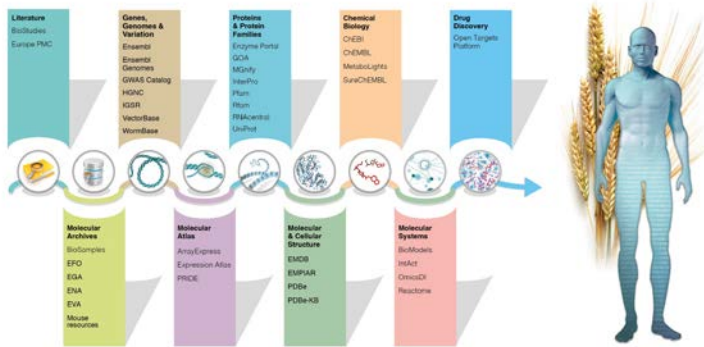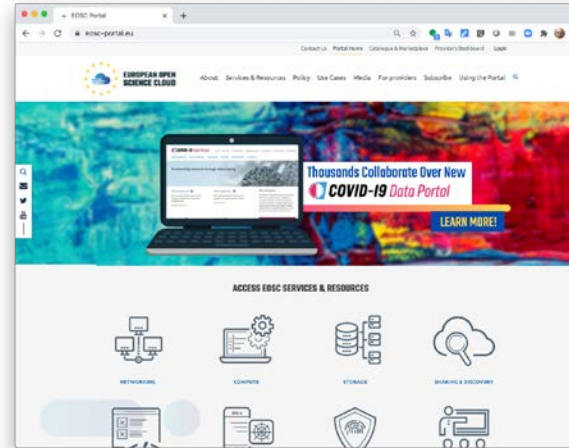
On 20 April 2020, Ursula von der Leyen, President of the European Commission, recorded a video message at the European Commission in Brussels, Belgium, on the launch of the EU COVID19 Data Platform.

On this platform, researchers will be able to store, exchange and analyse a wide range of knowledge about the

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

European Commission

Strong foundations

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

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

https://elixir-europe.org/
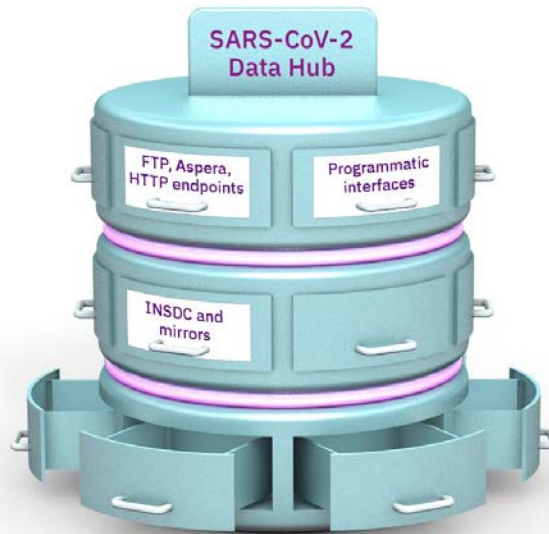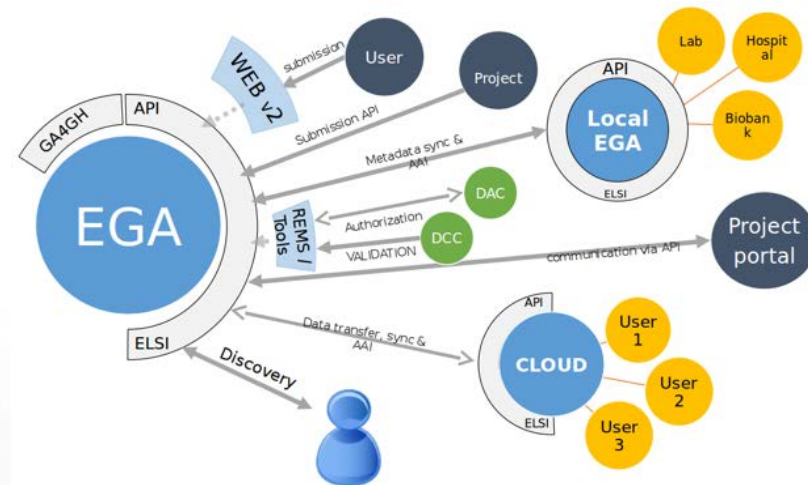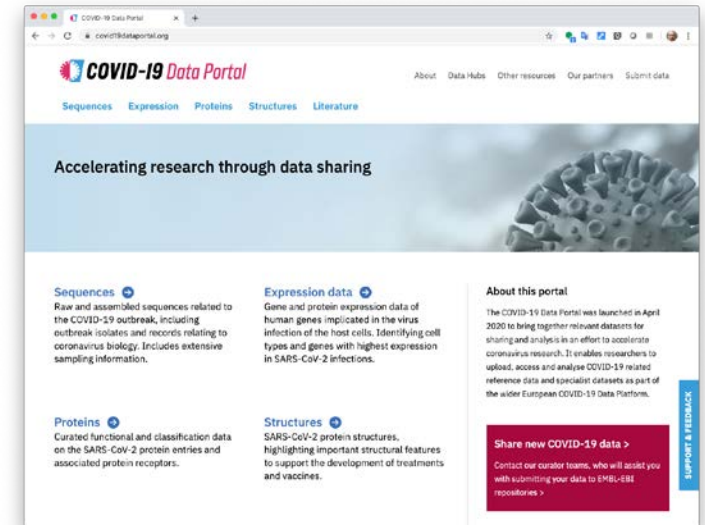
Core technical components

SARS-CoV-2 Data Hubs

Federated European
Genome-phenome Archive

COVID-19 Data Portal

https://www.covid19dataportal.org/

Data flows
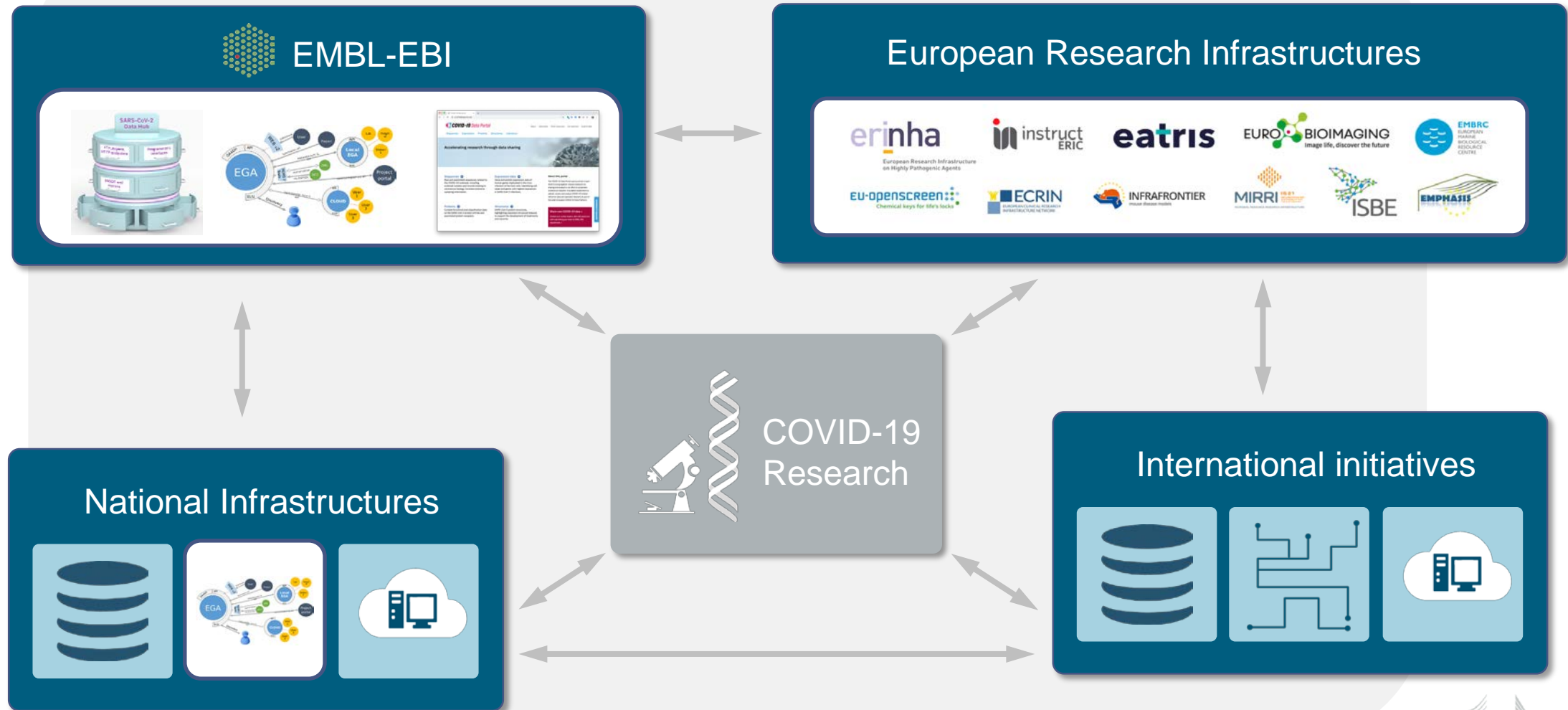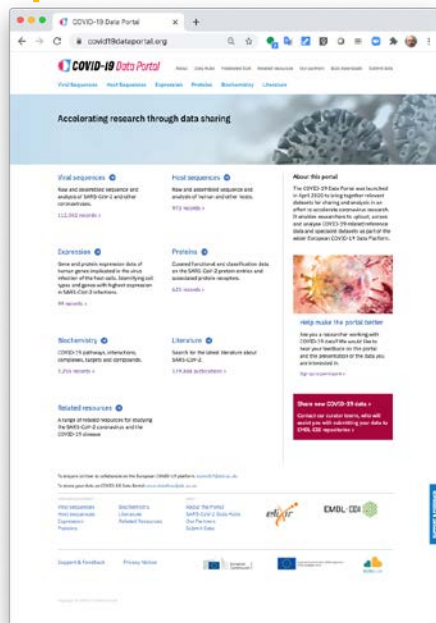

academia
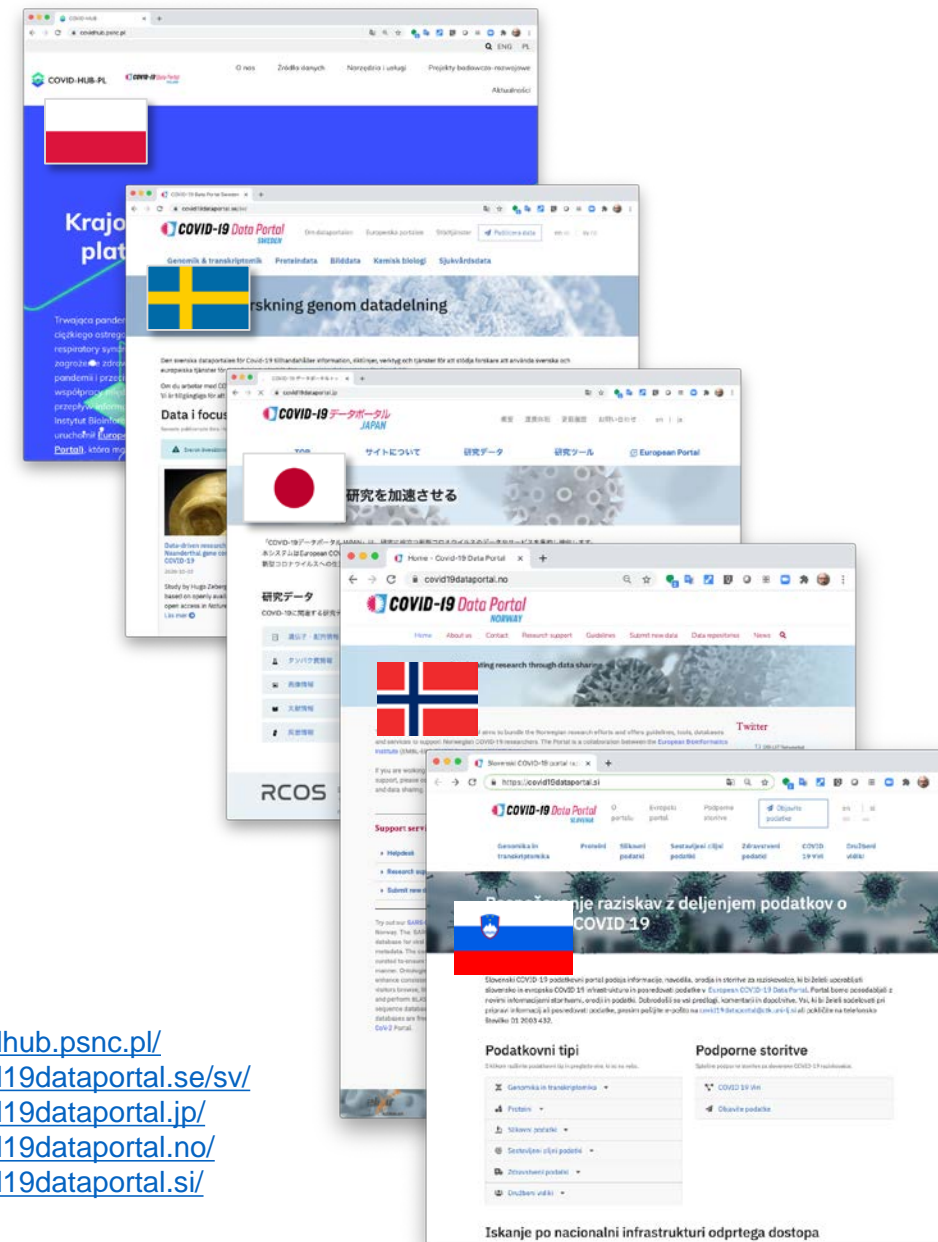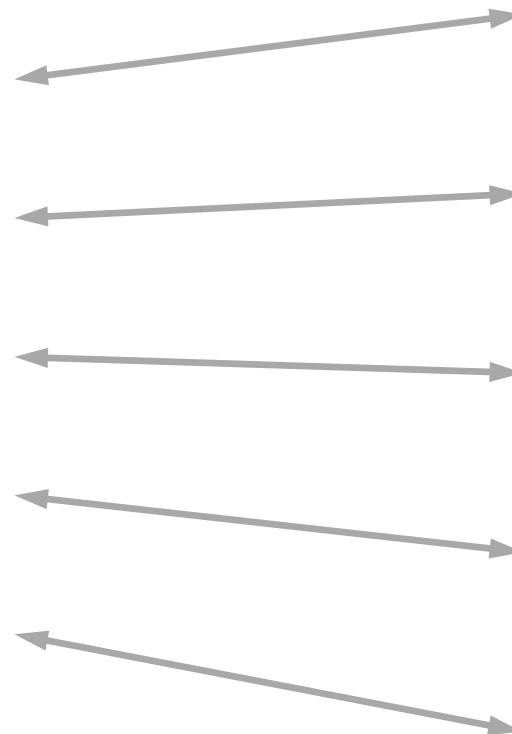
government

health

industry
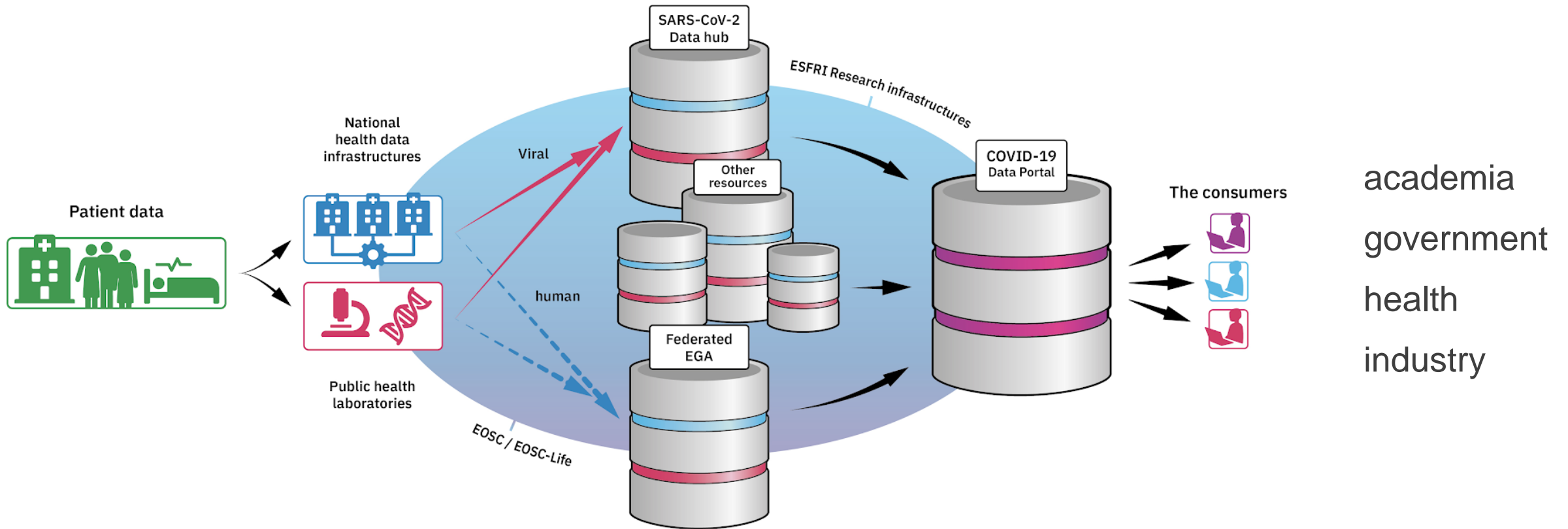
# Platform breadth

National coordination

Web toolkit

- National COVID-19 Data Portals to bring together national activities and users

- Technical support to implement and connect to Platform services

European Commission

National coordination

- Network of European national coordination teams activated by the EOSC Governance Board

- Scientific, bioinformatics, public health, health, political, funding

- 5 national COVID-19 Data Portals

- 16 national SARS-CoV-2 Data Hubs

- >5 emerging national nodes under Federated EGA
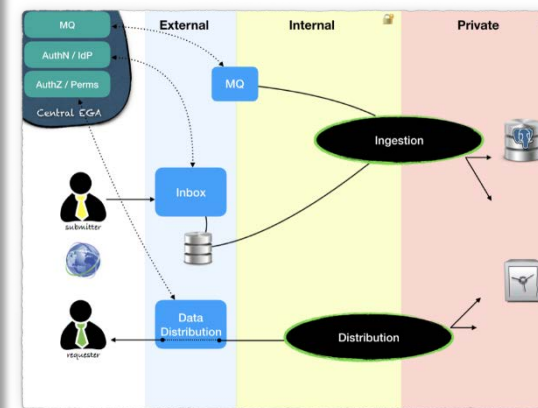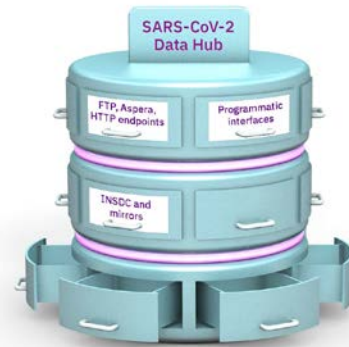
FAIR data and Open practices

http://www.insdc.org/

- Open data

- Data protection for sensitive human data

- Open software

- Open data standards

https://www.ga4gh.org/

https://beacon-network.org/#/

https://gensc.org/

https://www.globalmicrobialidentifier.org/
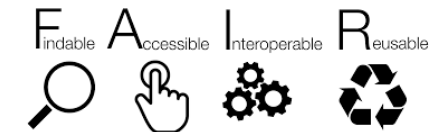

Findable Accessible Interoperable Reusable

Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016 Mar;3:160018. DOI: 10.1038/sdata.2016.18.

Data integration work stream

Viral Sequences	Host Sequences	Expression	Proteins	Biochemistry	Literature
All (262,317)	All (1,046)	All (82)	All (712)	All (2,255)	All (212,276)
Sequences (27,649)	Human studies (controlled access) (4)	Gene expression (4)	Protein sequences (57)	Pathways (16)	Coronaviruses (109,172)
Raw reads (125,159)	Human reads (consented for full access) (784)	Gene expression experiments (23)	Protein families (126)	Interactions (1,812)	Diseases (96,222)
Sequenced samples (96,632)	Other species reads (250)	Single cell expression (4)	Protein structures - Knowledge Base (7)	Complexes (29)	Related viruses and diseases (2,109)
Studies (163)	Association studies (8)	Single cell expression experiments (19)	Protein structures (260)	Compound document (8)	Genes, receptors and antibodies (4,773)
Genes (22)		Protein expression experiments (32)	Electron microscopy density maps (250)	Drug targets (390)	
Browser (1)			Electron microscopy public image archive (12)		
Variants (12,691)					

SARS-CoV-2
Genome submitted
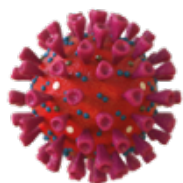
Launch of COVID-19 Data platform

ENA
EMDB
EPMC
Expression Atlas
InterPro
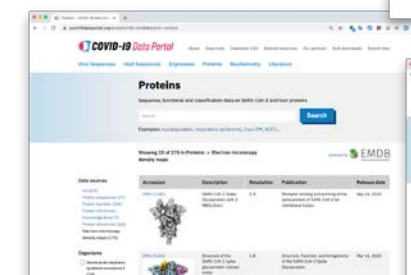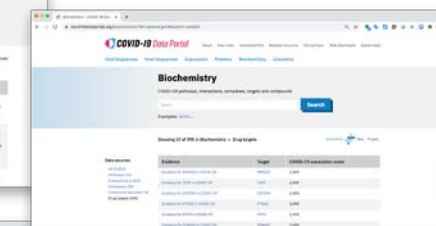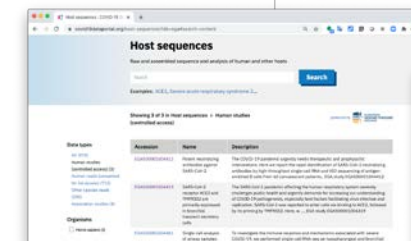PDBe
PRIDE
UniProt

ChEMBL

Reactome
Complex Portal
Ensembl

6 month anniversary

Data submissions/growth by EMBL-EBI services,

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov

European Commission

Viral genetic variation work stream

Viral data mobilisation

- Tools and services to support data upload

- Extensive support desk and campaign

- Standards and compliance support

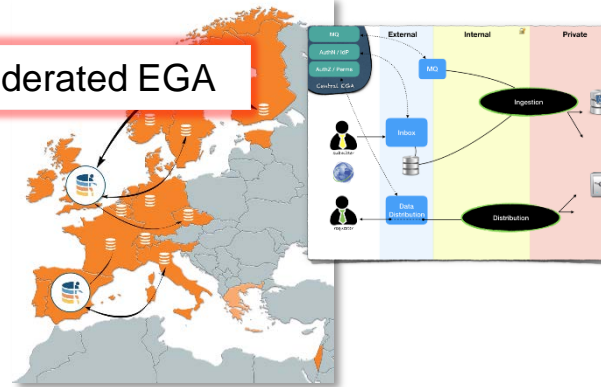- Curation, harmonization and quality control

Results to date

- Over 125,000 sequenced isolates from 47 countries

- 16 national SARS-CoV-2 Data Hubs supporting 74% of data flow

- Systematic analysis providing assemblies and genetic variations
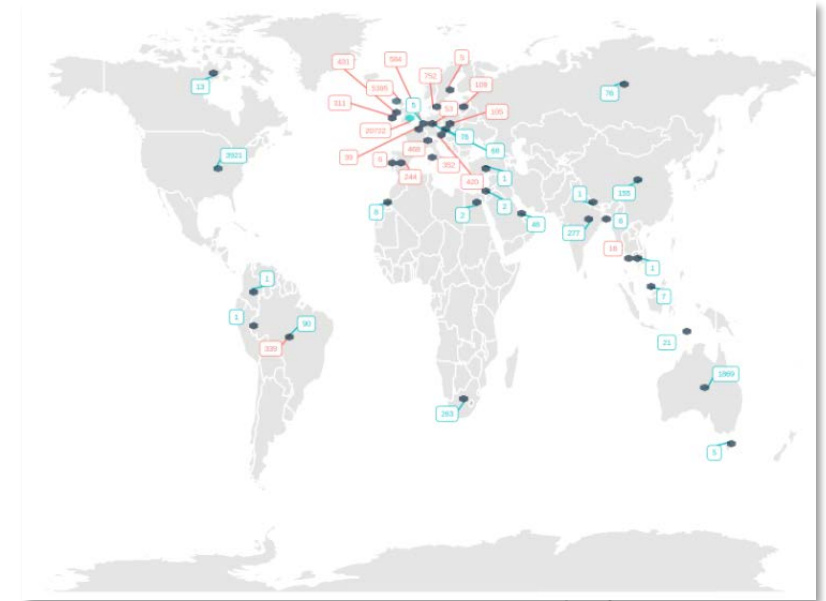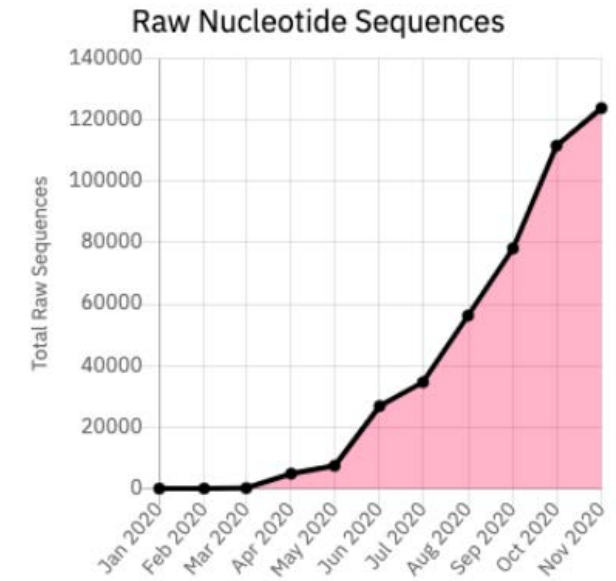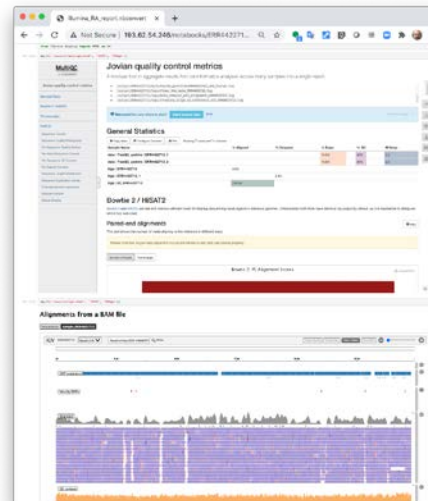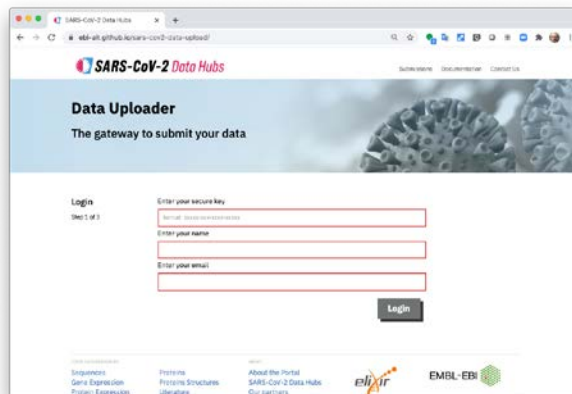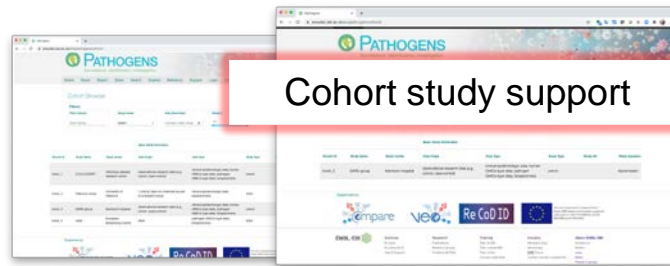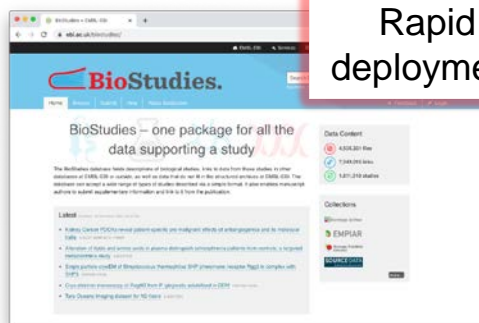
Example work streams
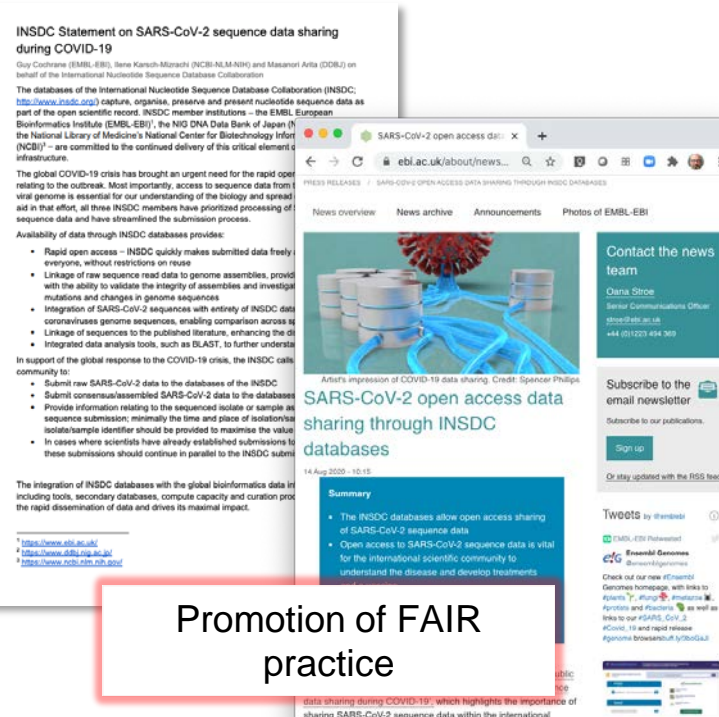

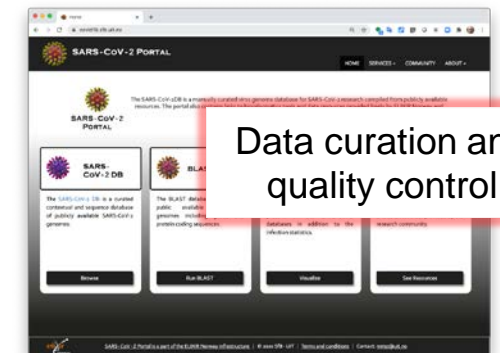Federated EGA

Analysis tool development

Cohort study support

Rapid deployment

Promotion of FAIR practice

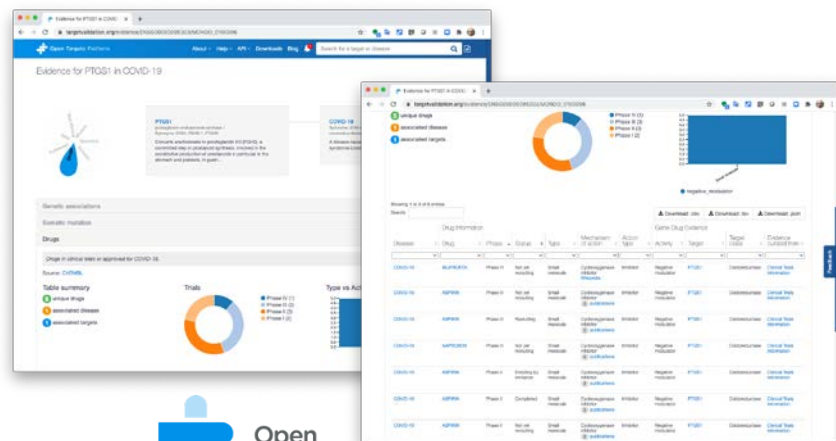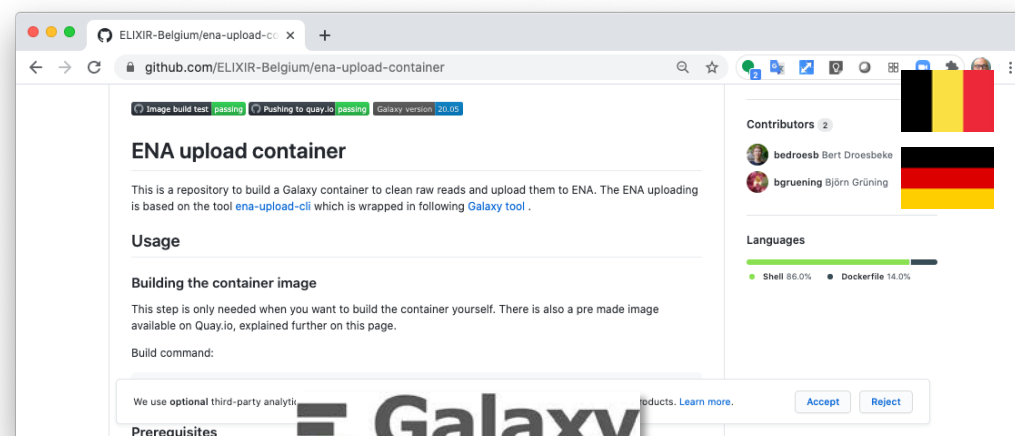Data curation and quality control

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

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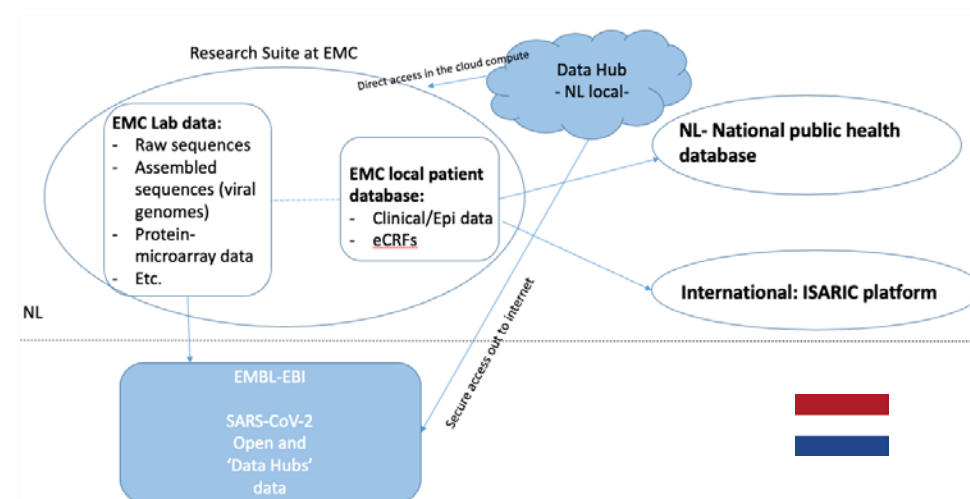
Use cases

Drug development

Connecting clinical/epidemiological data

Informatics

New tool simplifies the submission of SARS-CoV-2 data to open databases

ELIXIR Belgium and ELIXIR Germany (de.NBI) help researchers share FAIR COVID-19 data

ELIXIR Belgium, in collaboration with ELIXIR Germany and the European COVID-19 Data Platform, have developed a tool to simplify the submission of viral sequencing data to the European Nucleotide Archive (ENA), an ELIXIR Core Data Resource providing open access to nucleotide sequences. The new submission tool offers an easy-to-use interface, guides researchers through the submission process and verifies the data format and description.


Why submit data to ENA?

The research community has undertaken an unprecedented effort to study the novel coronavirus and its disease, COVID-19. Thanks to this coordinated effort, there is now an abundance of SARS-CoV-2 genomes freely accessible through multiple international databases.

However, some databases, such as GISAID, overlook the underlying raw data and only provide controlled access to consensus pre-assembled genomes. ENA provides a clear advantage to this restriction — it gives **access to both assemblies and raw data**. The latter is not only essential to reproduce findings but also to repurpose data to address new scientific questions.

Next steps

- Immediate rapid extension of the Platform – tools, data, connections, support, etc.
- Current discussions for linking with clinical, epidemiological and SSH data
- What we have learnt about collaborating around infectious disease
  - Open data and broad collaboration
  - Layers of communication: from national to global
  - Use, reuse, extension and adaptation of existing assets
  - Technical standards: simple, flexible and community-led
- Relevance to the next outbreak, surveillance and health research in general
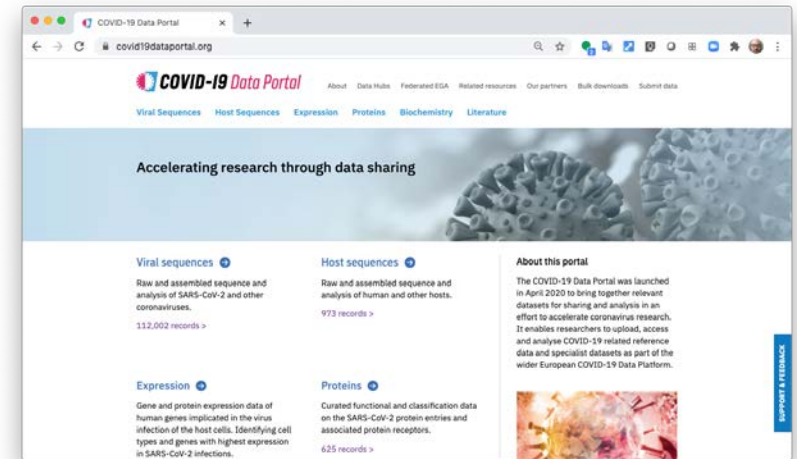
Thank you

If you reuse any slides please credit Guy Cochrane at EMBL-EBI