EMBL-EBI

The European COVID-19 Data Platform:

open science for human health

The European COVID-19 Data Platform has been instrumental in mobilising open access to SARS-CoV-2 data across the globe. EMBL's European Bioinformatics Institute (EMBL-EBI) and partners have coordinated national COVID-19 activities to connect efforts and data under one umbrella to spur scientific discovery, novel therapeutics and vaccines. This rapid response is only made possible through long term sustained funding of foundational open data resources.

IMPACT CASE STUDY

COVID-19 Data Platform

www.covid19dataportal.org/

Scientific response

In early 2020, as COVID-19 spread rapidly from country to country, with growing case numbers and rising mortality rates, **scientists, industry and health services** raced to understand this enormous, emergent threat to humanity.

In this response, the means for rapid data sharing soon became critical in understanding the biology, epidemiology, transmission, and evolution of the SARS-CoV-2 virus. The urgent need for transparent, secure and scalable sharing of pathogen data was key for the ongoing global management of this new infectious disease.

Accelerating research through data sharing

To meet this need, in April 2020 EMBL-EBI launched the European COVID-19 Data Platform in conjunction with the European Commission, the European Open Science Cloud (EOSC), ELIXIR and other partners.

The European COVID-19 Data Platform makes high-quality SARS-CoV-2 genomic sequence data and analysis tools freely available to the global scientific community.

As a trustworthy provider of open biological data for many decades, EMBL-EBI's rapid response was built on a rich history of large-scale international scientific collaborations, sustained investment in research data infrastructure, and direct host nation capital investment in EMBL-EBI's technology, through UK Research and Innovation.

"The European COVID-19 Data Platform has brought the European stakeholders together with one common goal: providing open and rapid access to data, tools and workflows."

Christine Durinx, Co-director, Swiss Institute of Bioinformatics

"I truly appreciate the importance of the European COVID-19 Data Platform as the umbrella for the integration of national data, but also its potential as a role model in the European health space."

Alfonso Valencia, Director, Spanish National Bioinformatics Institute

Value for the scientific community and the public

The European COVID-19 Data Platform has created value in many forms:

- **Mobilised open data**
  The Platform tools for data deposition have driven the direct submission of 85% of global viral raw genomic sequence data into the public domain via the Platform's Data Hubs. This includes the UK surveillance data from COG-UK, currently the largest SARS-CoV-2 dataset from a single organisation.

- **Ease of access to multiple, integrated sources of data**
  Provided integrated data and views to other SARS-CoV-2 related information, such as protein structures, compounds, drug candidates, as well as instruments to support users, such as compute infrastructure, workflows, and analysis tools.

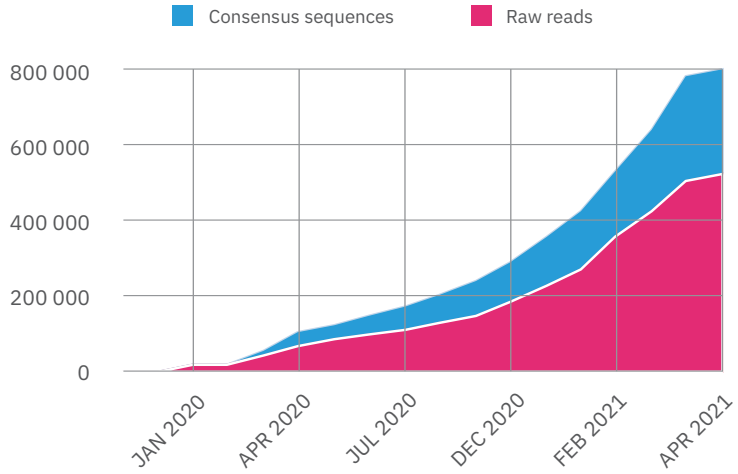
- **European coordination**
  The Platform provided a practical recipe for countries on how to coordinate national SARS-CoV-2 data through toolkits and EMBL-EBI technical support to implement their own National Data Portals.

- **Centralised data**
  The availability of both patient and research data from around the globe in one location enabled efficient data sharing, access and comparative analyses by researchers from diverse disciplines.

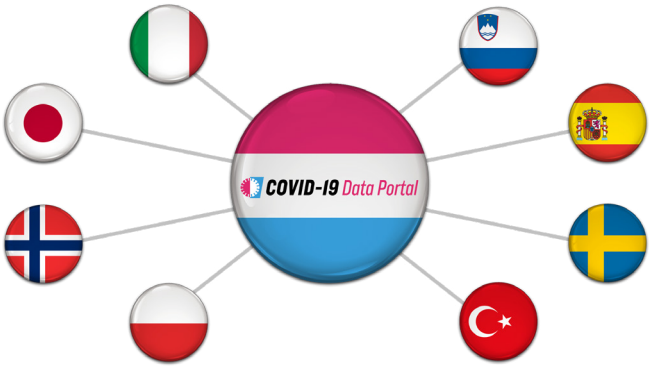
- **Connectivity**
  Facilitated access to up-to-date information for European member states on the progress of national COVID-19 efforts to collectively tackle shared challenges and benefit from shared solutions.

- **Data standards**
  Initiated the coordinated development of internationally-used minimum data standards for serology (diagnostic identification of antibodies in blood serum) and seroprevalence data.

Raw reads and consensus sequences



Open SARS-CoV-2 data increasingly mobilised through the European COVID-19 Data Platform.



Eight National Data Portals have been set up to support national coordination of SARS-CoV-2 efforts.

4 million
Web requests

177,000
Users

1.4 million
Data records

COVID-19 Data Portal usage between April 2020 and April 2021.

Economic impact

EMBL-EBI's forthcoming economic impact study has also shown that the institute's data resources, which underlie the Platform, have been making a major contribution to science throughout the pandemic. Of the 4,000 survey respondents who answered the COVID-19 questions, 32% said they valued EMBL-EBI data resources more as a result of the pandemic, and 25% said they had used them more; 18% of respondents were users of the European COVID-19 Data Platform, a relatively significant proportion for a newly-established data resource.

The responses to the 2021 economic impact survey underscore the efficiencies that open data services provide to researchers in tackling global challenges. In 2016 overall efficiency of EMBL-EBI resources was conservatively estimated by users at £1-5 billion per annum. Higher efficiencies are observed in the current assessment responses.

"This is a breakthrough in open science, expressed in the determination to share genetic sequence data of the virus."

Cezary Mazurek, Director, Poznań Supercomputing and Networking Center

"Losing access to EMBL-EBI services and resources would constitute a major setback for infection control and research."

2021 EMBL-EBI user survey respondent

Preparing for future pandemics

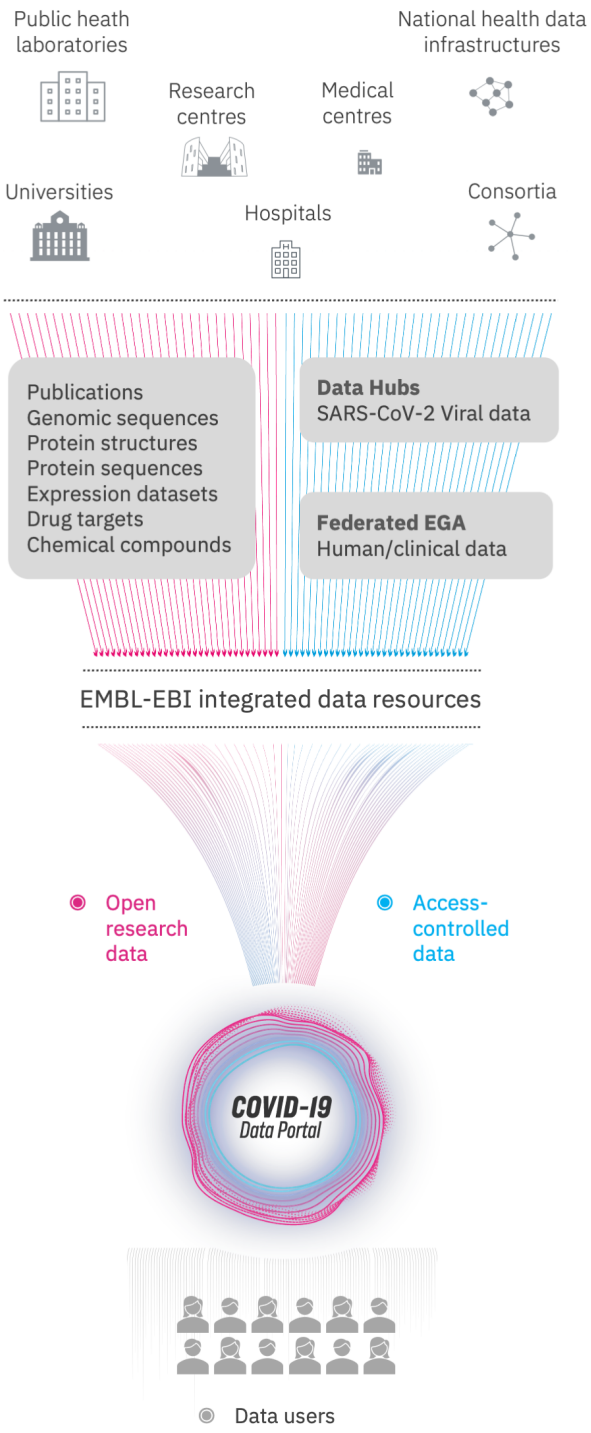
The current pandemic is far from over and continued work is required to support both open SARS-CoV-2 data mobilisation, and integration of even more data types to the Platform, such as epidemiology, population health outcomes, economic and social research data. Ensuring the sustainability of EMBL-EBI's integrated biological data resources will provide durable core services for Europe and the world in times of both peace and crisis.

Preparations for future pandemics highlight how imperative it is that countries pool resources, cooperate, and align strategic directions to drive forward scientific progress and to deliver research that is relevant to pressing societal challenges.

Platform components

The European COVID-19 Data Platform uniquely brings together three interconnected components that provide easy and secure access to patient and research data:

- Data Hubs can be used to organise SARS-CoV-2 genomic sequence data at the national, regional and institute levels. Upon release, this data can then be openly shared with the global research community via the COVID-19 Data Portal.
- The Federated European Genome-phenome Archive provides secure, controlled-access sharing of sensitive COVID-19-related human genomic and phenotype data from patients.
- The COVID-19 Data Portal brings together tools and SARS-CoV-2 viral and host datasets spanning genomics, proteins, imaging, drug compounds and publications, enabling researchers to access and analyse integrated COVID-19 data easily.



Flow of SARS-CoV-2 patient and research data into the COVID-19 Data Portal.

Explore further

COVID-19 Data Portal
covid19dataportal.org

EMBL-EBI open letter: Support data sharing for COVID-19
www.covid19dataportal.org/support-data-sharing-covid19