

The European COVID-19 Research Data Platform

1. Context and objective

The coronavirus crisis is putting high pressure on the research community to speed up science discovery, inform public health policy and respond to the crisis. Rapid data sharing plays a key role in accelerating the understanding of the biology, epidemiology, transmission, and evolution of the virus.

The European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory (EMBL) and, the Commission, together with other partners¹ have recognised the urgency to develop and deploy a pan European COVID-19 research data platform connected to the European Open Science Cloud (EOSC).

The objective is to speed up and improve, within the shortest timeframe, the sharing, storage, processing of and access to research data and metadata on the SARS-CoV-2 and COVID-19 disease. The initiative will start giving access to EMBL-EBI sequence data from mid-April 2020 and progressively expand its data, workflows and processing offering over the next two years. With the help of the Member States and other stakeholders, widest possible dissemination of this platform will be ensured.

The initiative builds notably upon existing connectivity between EMBL-EBI infrastructure and national public health data infrastructures. It is in line with the overall objectives of EOSC, complements the relevant activities by the European Research Infrastructures, ongoing research projects as well as actions by international and national research organisations. The initiative is central to a wider Commission plan on Covid-19 related actions in the field of Research and Innovation² supporting the European response to the crisis.

2. Link to EOSC and the Governance Board

The deployment of a European research data platform for sharing COVID-19 data is seen as a priority thematic pilot initiative to realise the EOSC vision of creating a European data space for science, research and innovation. Given the unprecedented number of research initiatives helping to combat the global Covid-19 pandemic, the EOSC Governance Board provides an immediate formal structure to address the coordination challenge. This Board shall be used as the main structure to interact with the Member States and Associated Countries on the European COVID-19 Data Platform.

3. Concept for the intended data platform

The European Research COVID-19 Data Platform will provide an open, trusted, and scalable pan European environment where researchers can store and share relevant datasets including:

- ✓ Omics data for the characterisation and quantification of biological molecules (including sequence data on both virus genomes and human genomes) and other high-dimensional data such as microbiome data;
- ✓ Data from pre-clinical research to test drug candidates, vaccine interventions, or other treatments, for efficacy, toxicity and pharmacokinetic information;
- ✓ Research data from clinical trials and from observational studies;
- ✓ Epidemiological data, models, codes and algorithms.

It will also provide systems for data exploration and visualisation and a cloud compute facility where scientists and public health workers can collaborate.

¹ Including ELIXIR, Instruct-ERIC, RDA and H2020 projects such as EOSC-Life, CORBEL, RECODID, VEO, EXSCALATE4CoV.

² Including complementary actions to (i) coordinate and connect to the EOSC data services by all relevant European Research Infrastructures, (ii) leverage information available in European databases and registries on clinical trials and observational research, (iii) use high performance computers for data analysis, (iv) develop specific Covid-19 data management guidelines, (v) develop rapid and coherent data provisions from relevant Horizon 2020 projects and (vi) facilitate access to data underlying existing and forthcoming Covid-19 publications.

The European COVID-19 Research Data Platform³ consists of two connected components. The COVID-19 Portal will be the main interface for the researchers, bringing together and continuously updating relevant COVID-19 datasets and tools. The SARS-CoV-2 Data Hubs will organise the flow of research data from the outbreak and feed the COVID-19 Portal. Essential metadata will be captured from the various Data Hub which will differ to reflect national and regional efforts and requirements (see annex).

4. Data sharing and management

As a principle, all data and metadata accessible from the Covid-19 Portal will be open and as FAIR⁴ as possible. Existing standards already used by the research community will be favoured. In the absence of community standards, specific guidelines will be developed that will adhere to the FAIR principles and will take into account the specificities related to omics data, pre-clinical and clinical research data and epidemiological data. These guidelines will be developed in the context of the new RDA COVID-19 Working Group in synergy with existing initiatives such as the Virus Outbreak Data Network (VODAN)⁵ and the GloPID-R alliance⁶. EU data legislation will be respected in particular personal data protection.

5. Data deposition

EMBL-EBI, in consultation with other partners, will develop specific guidelines for data deposition in the platform over the coming two weeks. They will also launch a survey, targeting Horizon 2020 projects at first, to gather requirements on depositing different data types.

6. Link to other COVID-19 related platforms

The European COVID-19 Research Data Platform will link with information exchange or communication platforms targeting diverse audiences including the clinicians and public health authorities (e.g. the Health Policy Platform operated by the Commission or platforms operated by WHO). It will also link with other European research platforms and supercomputing activities such as the PRACE call for COVID-19 or the exploitation of the Exscalate platform for COVID-19⁷.

7. Contacts

- For Covid-19 research data sharing policy and EOSC GB matters:
Thomas Neidenmark (Thomas.Neidenmark@ec.europa.eu)
- For questions related to the European Covid-19 Research Data Platform:
Dr Rolf Apweiler, Director, EMBL-EBI (apweiler@ebi.ac.uk)

In Annex: Description of the components of the European COVID-19 Research Data Platform (separate document)

³ www.ebi.ac.uk/covid-19

⁴ Findable, Accessible, Interoperable, Reusable

⁵ www.go-fair.org/implementation-networks/overview/vodan/

⁶ www.glopid-r.org/

⁷ www.exscalate4cov.eu/