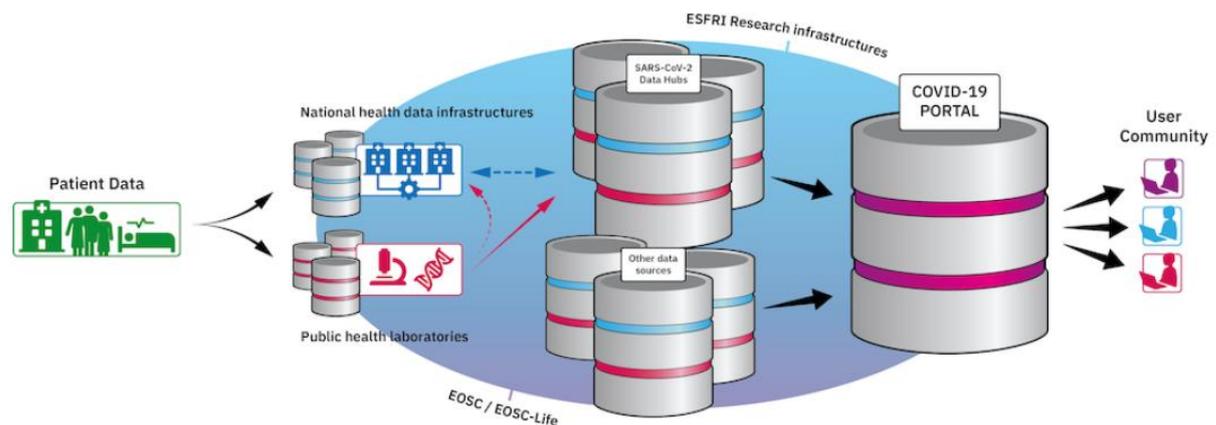## <u>Annex</u>: Components of the European COVID-19 Research Data Platform

The European COVID-19 Research Data Platform[1] consists of two connected components: the COVID-19 Portal (which is the main interface to the research community) fed by the SARS-CoV-2 Data Hubs:

- <u>The COVID-19 Portal</u> will be the main interface for the researchers, bringing together and continuously updating relevant COVID-19 datasets and tools. In a first stage, the COVID-19 Portal will feature relevant datasets from EMBL-EBI data resources such as the European Nucleotide Archive (ENA), UniProt, Protein Data Bank in Europe (PDBe), the Electron Microscopy Data Bank (EMDB), Expression Atlas, and Europe PMC. The portal will also include the outbreak sequence data and a Cohort Browser for searching clinical and epidemiological data (including by means of a metadata catalogue). It will also enable scientists to upload, search, and explore specialist datasets. In a second stage, additional datasets and tools from other European projects / existing platforms will be accessible with the long-term objective of including data from other international projects and European research infrstructures. This will be achieved with the help of the Commission, the EOSC governance and ELIXIR, the intergovernmental organisation that brings together life science data and resources from across Europe, and other collaborators.

- <u>The SARS-CoV-2 Data Hubs</u> will organise the flow of research data from the outbreak and provide comprehensive open data sharing for the research communities, starting with genomics data and expanding toother types of data. It will build on the EMBL-EBI infrastructure and will mainly be used by scientists (and public health agencies) responsible for generating viral sequences, microbiome data, data on host genetics and immune response or epidemiological modelling at national or regional levels. The research data in each Data Hub will differ to reflect national and regional efforts and requirements. Essential metadata will be captured, including sampling time, method, geographical location, sequencing technology, and the health status of the host. The Data Hubs will also provide systematic data processing, visualisation, and phylogenetic analysis tools.

Source: EMBL-EBI[2]