



CONCEPT NOTE

Open Science and the European COVID-19 Research Data Platform

Introduction

The COVID-19 pandemic is putting high pressure on the research community to accelerate science discovery, inform the public health response and help save lives. To accelerate research and amplify its impact we need to ensure that research findings and data relevant to fighting the pandemic are shared as rapidly, openly and effectively as possible.

Open Science policies and practices are best suited to foster an environment where the sharing of knowledge, data and tools happen as early as possible, not only between researchers and between disciplines, but also with society at large. In addition, Open Science improves the quality, efficiency and creativity of research, while increasing society's trust in science.

The European Open Science Cloud (EOSC), an Open Science flagship, responds to high demand among the research community, amplified by the pandemic, to federate existing and emerging infrastructures and to offer a trusted, open and distributed ecosystem that provides seamless access to data and interoperable services addressing the whole research data lifecycle.

The European Commission and EMBL's European Bioinformatics Institute ([EMBL-EBI](#)), together with [ELIXIR and other partners](#), have recognised the urgency to respond to data-sharing needs in research by creating and deploying a dedicated European COVID-19 research data platform, fully integrated within the European Open Science Cloud (EOSC). This joint effort to rapidly deploy a European COVID-19 research data platform is a priority thematic pilot initiative to realise the EOSC vision and is facilitating comprehensive data sharing for the European and global research communities.

As a short-term action, we invite researchers from Africa to access and reuse data currently available on the platform to accelerate their work on COVID-19, as well as to submit relevant data that can in turn be reused by others helping to fight the pandemic. We further propose to co-organise a webinar with EMBL-EBI/ELIXIR for interested African researchers to familiarise themselves with using the platform and its current and upcoming features, as well as to increase the FAIRness of all COVID-19 related datasets.

For medium - and long-term actions, we suggest to explore synergies and maximise alignment with the African Open Science Platform (AOSP), e.g. through a pilot project to streamline COVID-19 data-flows between the two platforms (in partnership with NRF South Africa). In addition, we invite relevant stakeholders from African nations dealing with clinical and epidemiological data to participate in our future meetings to discuss how we can make progress in sharing those types of data globally.

Background

The COVID-19 crisis has demonstrated the added value of addressing the rapid and open sharing of FAIR data as a way of increasing the impact of research. The European Commission has created its [ERAvsCORONA Action Plan](#) which was approved by the Directors General of the R&I ministries in the EU Member States. This plan contains 10 key areas for coordinated R&I actions.

One such action is the setting up of a research data sharing platform, under the umbrella of the European Open Science Cloud, to help improve the sharing, re-use, processing of and access to research data and metadata relevant to the current outbreak. The European COVID-19 research data platform is being developed by the European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory (EMBL), in partnership with the European Commission, and in collaboration with relevant partners, such as ELIXIR, the European Research Infrastructure for Life Sciences.

The platform will be crucial to speed up, and improve, within the shortest timeframe, the sharing, reuse, processing of, and access to, research data and metadata on the SARS-CoV-2 and COVID-19 disease, across research fields, and across national borders. As such, the platform represents a concrete European R&I response to demonstrate a thematic implementation of the European Open Science Cloud (EOSC) vision, i.e., an open, distributed and trusted ecosystem, through which researchers will be able to seamlessly access data and services addressing the whole research data lifecycle, from genomics and protein structures, to preclinical, clinical and epidemiological data.

The [COVID-19 data portal](#), an important entry point to the COVID-19 research data platform, was formally launched on 20 April by the European Commission and EMBL-EBI.

The platform's development is supported by topping-up the budget of existing Horizon 2020 projects under the Health and Research Infrastructures parts. This will allow either adding or re-orienting activities within these ongoing projects. The development of the platform illustrates a concerted effort between different parts of Horizon 2020 and policies, including the Open Science policy, together with the help of Member States. As a concrete example, Sweden recently launched their [National mirror of the COVID-19 data portal](#) and this sets an example to be followed by other Member States and Associated Countries, strengthening the idea of a federated platform that connects data across borders for the benefit of advancing the global understanding of how the novel coronavirus propagates and how the COVID-19 disease spreads.

The initial platform development was centered on access to omics data. Omics data indeed need to be shared as early as possible in order to facilitate the identification of genetic determinants of COVID-19 susceptibility, severity and outcomes. As of today, the portal offers access to viral sequences, host sequences, protein sequences, structures and expression data, as well as relevant scientific articles and preprints. More than 25.000 raw or assembled sequences are, for example, made available. Since 28 May, compounds activity data are also available, which is of vital importance for potential drug repurposing research. This will be broadened further by bringing in compounds data from the European research infrastructure [EU-OPENSREEN](#). Current efforts on the platform are now focused on incorporating additional data types. This concerns protein-protein interactions data, serolomics data, but most importantly clinical data. The sharing of clinical metadata and data in a suitable, trustworthy, and secure way is crucial considering the many clinical trials and studies performed under enormous time pressure. The [ECRIN research infrastructure](#) has launched a [Clinical Research Metadata Repository](#) that includes COVID-19 metadata and enables the

discovery of clinical studies and related data objects such as protocols, consent forms, data management plans, case report forms, descriptive metadata and more. The Metadata Repository is available from the portal and there are active discussions to improve this offering in the next release of the COVID-19 data portal. Discussions are also ongoing on how to link with metadata and data from cohorts. In order to link clinical data with genetic data, the platform will also build on the federated [European Genome-phenome Archive](#) (EGA) an archive that promotes the distribution and sharing of genetic and phenotypic data consented for specific approved uses.

Countries in Africa have started to invest in elements of policy at national level that facilitate open science. The African Open Science Platform (AOSP) strategy makes a case for open science in Africa for the benefits it can deliver to science, society and the economy. The AOSP Open Data Policy Framework provides guidelines for the implementation of open data policies, for consideration by governments, scientists and their institutions.

Other initiatives driven by sectors such as astronomy have also been undertaken, as it needs systems to deal with big data systems. For example, the Inter-University Institute for Data Intensive Astronomy in South Africa seeking to build capacity and expertise in data intensive research. The Ilifu data-intensive facility which has a focus of research and development for cloud-based data-intensive research solutions for the strategic science domains of astronomy and bioinformatics. In 2019, the Institution has also engaged with the EGI Foundation in Europe and has completed a demonstration federated cloud platform with EGI. The demonstration project enabled South African and European users to access ILIFU cloud services and infrastructure as well as resources on the EOSC.

The National Integrated Cyber Infrastructure System (NICIS) promotes scientific and industrial development through the provision of high-performance computing capability, high-speed network capacity and a national research data infrastructure integrated hierarchically into globally connected systems and into local system systems, providing seamless access for the research and education communities of South Africa. NICIS is also providing SADC Cyber-Infrastructure, and potentially could look at providing a mirror image of a COVID-19 data portal.

Objectives

Our immediate objective is to raise awareness of the COVID-19 research data platform and to update on the latest developments. We would also like to invite African researchers to both reuse data from and upload relevant COVID-19 data to the portal. We look forward to exploring ways to connect with the African Open Science Platform, as well as to link activities of the European and Developing Countries Clinical Trials Partnership (EDCTP) and improve sharing of data from clinical trials across Europe and Africa.

In another development, more than 130 scientific publishers and funders (of which the European Commission is one) have reacted to the COVID-19 crisis by signing the [Statement on Sharing research data and findings](#) relevant to the novel coronavirus outbreak, committing to promote immediate open access to publications, data and results, at least for the duration of the outbreak. We invite members of the HLPD bureau to consider signing the statement and taking active steps to demonstrate how Open Science practices can provide effective solutions against the current pandemic.

Funding

Potential sources of funding could include the [EOSC Secretariat](#): co-creation funding or COVID-19 fast track funding for pilot projects that can bring relevant data to the COVID-19 data portal and make data FAIR or create concrete links between the COVID-19 data platform and AOSP.

Training activities could be funded jointly with EMBL-EBI and/or ELIXIR.