

Data Together COVID-19 Appeal and Actions

March 2020



The COVID-19 pandemic presents a major test for our science system and for our research and data infrastructures. These infrastructures, such as open science clouds and data commons, must serve the needs of science, policy, and humanity not only in ‘normal times’, but also in times of crisis by providing controlled access to quality data in real time and at scale for a range of scientific- and policy-related responses.

Governments and funders across the globe are launching and funding initiatives to ensure that pivotal decisions are made based on the best available science to mobilise stakeholders for the current situation.

There is the need and opportunity to unite existing initiatives and accelerate the development of core services to help meet the current crisis. We must manage expectations and be clear on the limitations of what can be achieved. At the same time, it is essential to rise to the challenge and ensure that what is implemented for COVID-19 is sustainable and scalable.

The Data Together Organisations—**comprising CODATA, GO FAIR, RDA and WDS**—in response to the problems listed below, jointly contend that it is essential to meet both the immediate needs and the long-term objectives of global science by ensuring that data and science platforms and infrastructures are based on the FAIR Principles.¹ This will maximise the ability to combine, visualise, and use data from many sources; facilitate fine-grained data access and protection; and allow for decentralised and machine-assisted analysis.

In summary, the current situation requires the accelerated implementation of a FAIR ecosystem.² The need to act quickly should not lead to lowering our ambition as to what research and data infrastructure is needed, or to our accepting the view that a conventional portal is sufficient without an underlying FAIR infrastructure.

¹ Wilkinson, M., et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

² European Commission Expert Group on FAIR Data, *Turning FAIR into Reality*, (2018) <https://doi.org/10.2777/1524>

The COVID-19 platforms we need

The boundary considerations for an open science and FAIR data platform focussed on COVID-19 are:

- Some COVID-19 related data have political or institutional sensitivities, many have personal components. Access to sensitive and personal data must be restricted in various ways, and the institutions that manage these datasets cannot release them openly without additional processing or access controls. Such data can generally only be accessed partially and in controlled circumstances, and for this to be achieved, the data must be FAIR. We need, therefore, to make a rigorous effort in favour of FAIR, while continuing to emphasise the policy: *As open as possible, as closed as necessary*.
- A centralised, data warehousing approach is not possible and not fit for purpose. As the data are distributed, rich machine-actionable metadata are necessary to enable controlled, computational access for analysis or visualisation.
- Very large quantities of data are being generated in relation to the pandemic. There are significant challenges in ensuring data quality and risks of false and misleading information being disseminated as ‘fact’. There should be a mechanism to mitigate such dangers.
- We need to facilitate and further enhance methods for distributed deep learning, and make sure the algorithms and services on which such approaches are based can work effectively with FAIR metadata and, where possible, with FAIR data.
- We need a community annotation system that enables objective assessment of new claims and information. Decision makers can then rely on a vast community of trusted experts to review new and existing claims relevant for COVID-19 interventions.
- No particular public or private organisation should be able to monopolise the applications or the FAIR ecosystem. Therefore, a quality control and a minimal certification scheme for all components must be in place as part of the effort.

The Data Together organisations are jointly working on the following activities and invite all to contribute:

1. Review and support the [Data Together–GO FAIR Virus Outbreak Data Network](#) (VODAN), and if in agreement with the approach, join as a contributor or an observer endorsing this network. The primary objective of the VODAN Implementation Network is to showcase the creation and deployment of FAIR data related to COVID-19.
2. Review and support the [Data Together–RDA COVID-19 Working Group](#) that has the remit to define detailed guidelines on data sharing and reuse under the present COVID-19 circumstances, and thus help researchers follow best

practices to maximise the efficiency of their work, including the deposit of different data sources in any common data hub or platform.

3. CODATA has been tasked by the International Science Council to prepare a major global programme '[Making Data Work for Cross-Domain Grand Challenges](#)' for launch in 2021 with the Data Together organisations and a large number of other partners. The aim of the initiative is to assist interdisciplinary research by addressing the critical issues of data access, interoperability and reuse across domain, disciplinary, and institutional boundaries. Initial working groups are underway. In the area of infectious disease, work is focussing on combining large-scale longitudinal cohort data and clinical data in a secure and federated infrastructure for analysis of HIV. This work will be expanded to include an examination of pandemic epidemiology data, with reference to the urgent case of COVID-19, and in preparation for future resilience.

