# Enhancing access to research data during crises: lessons learned from the COVID-19 pandemic.

This is a background paper that has been prepared for an OECD Global Science Forum (GSF) workshop on 23 April, 2021, which is part of a broader project on *Mobilising science in response to COVID-19: lessons learned from COVID-19.*

This workshop is being organised with the Research Data Alliance (RDA) and is recognised as a co-located event for the 17 RDA Plenary meeting.

See [oe.cd/RDAworkshop](oe.cd/RDAworkshop)

A version of this document is also available to GSF delegates in O.N.E Members &Partners under the reference code DSTI/STP/GSF(2021)7

# Background Paper

# Enhancing access to research data during crises: lessons learned from the COVID-19 pandemic.

***Readers'note:*** *This draft background paper has been prepared for an OECD GSF workshop on 23 April, 2021, which is part of a broader project on Mobilising science in response to COVID-19: lessons learned from COVID-19. The paper will be amended and completed after the workshop*

## Covid-19 and data

Data access and re-use is critical for ensuring the efficiency and effectiveness of the global scientific research effort to address the COVID-19 pandemic. Access to trusted, well described data, together with the software, models and workflows that are necessary for the production and analysis of this data, is necessary across the many different domains of science that are being mobilised to understand and combat COVID-19. Transparent and timely access to research data and models underpins the science advice that is informing public health strategies. Such data and analytical tools have been critical to the rapid development and testing of vaccines and therapeutics. They are also essential for designing, monitoring and assessing the impact of the socio-economic policies that are being implemented in response to the pandemic.

The availability of information and communication technology has improved the global capacity to implement systems to share data during a pandemic. However, the lack of harmonisation across diverse systems is currently a major obstacle to timely and effective access. Initiatives already underway in Europe and other regions to develop Open Science Clouds are not yet well enough developed to overcome this obstacle. In short, a lot of data that are extremely valuable for COVID-19 research and responding to the pandemic are not sufficiently, findable, accessible, interoperable and reusable (FAIR).

The unprecedented spread of the virus has prompted a rapid and massive research response and this has been greatly facilitated by well-established international data sharing initiatives - such as GISAID for SARS-CoV-2 genomic data. However, such initiatives remain restricted to certain research domains and in many fields there are no universally adopted systems or standards, for collecting, documenting and disseminating COVID-19 research data and associated code and software. Many data are not reusable by, or useful to, different communities if they have not been sufficiently documented and contextualised or appropriately licensed. This is not a new challenge for many areas of science but, in the context of COVID-19, it is a challenge that needs to be urgently addressed and for which, in many cases, solutions exist but have not been fully adopted.

The responsible, FAIR and timely sharing of data is an essential element of the Open Science approach that the world needs to effectively combat pandemics like COVID-19 and other complex crises. Unnecessarily limiting or delaying access not only slows-up scientific progress but, as we are seeing in several countries, it can also undermine public trust in science and science-based decision-making. At the same time, much of the data related to the pandemic is personal or sensitive, and it is critically important to protect privacy and ensure the ethical management and use of this resource. Policymakers, research funders, and research institutions around the world must work together, to mandate, incentivise and support actions to harmonise and streamline the responsible and timely provision and exchange of data locally, nationally and internationally.

# OECD recommendation on enhanced access to research data

The OECD Council approved a revised [Recommendation on Access to Research Data from Public Funding](#) in January 2021 (hereafter the Research Data recommendation") as the World entered the second year of the COVID pandemic. The revised recommendation reaffirms the relevance and importance of several key principles that were originally set out in 2006: openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, and sustainability. In addition, it provides updated policy guidance on seven areas that have emerged as crucial for enhancing access to research data in recent years: 1. Data governance for trust; 2. Technical standards and practices; 3. Incentives and rewards; 4. Responsibility, ownership and stewardship; 5. Sustainable infrastructures; 6. Human capital; and, 7. International cooperation on data access. Importantly, the scope of the recommendation has been extended to include access to algorithms, workflows, models and software (including code).

Although the 2021 recommendation focuses on enhancing access to data from publically funded research, it also acknowledges that making all data fully open is neither feasible or desirable. It recognises that "access to some data will need to be partially or totally restricted to conform to legal rights, ethical principles and/or to protect legitimate private, public, or community interests." This 'open as possible, closed as necessary' approach is particularly relevant in relation to health pandemics, where much of the data that is required for research is personal or otherwise sensitive data.

The 2021 recommendation does not specifically address the COVID-19 pandemic situation or crises but its revision and release is timely as it provides a broad policy framework for enhancing data access that is applicable to crises. In an OECD Policy Brief [Why open science is critical to combatting COVID-19](#) (June 2020) it was noted that, while global sharing of research data has reached unprecedented levels, challenges remain. Outstanding issues that were highlighted, include the lack of specific standards, coordination and interoperability, as well as data quality assurance.

# Guidelines on data sharing in crises, including COVID-19
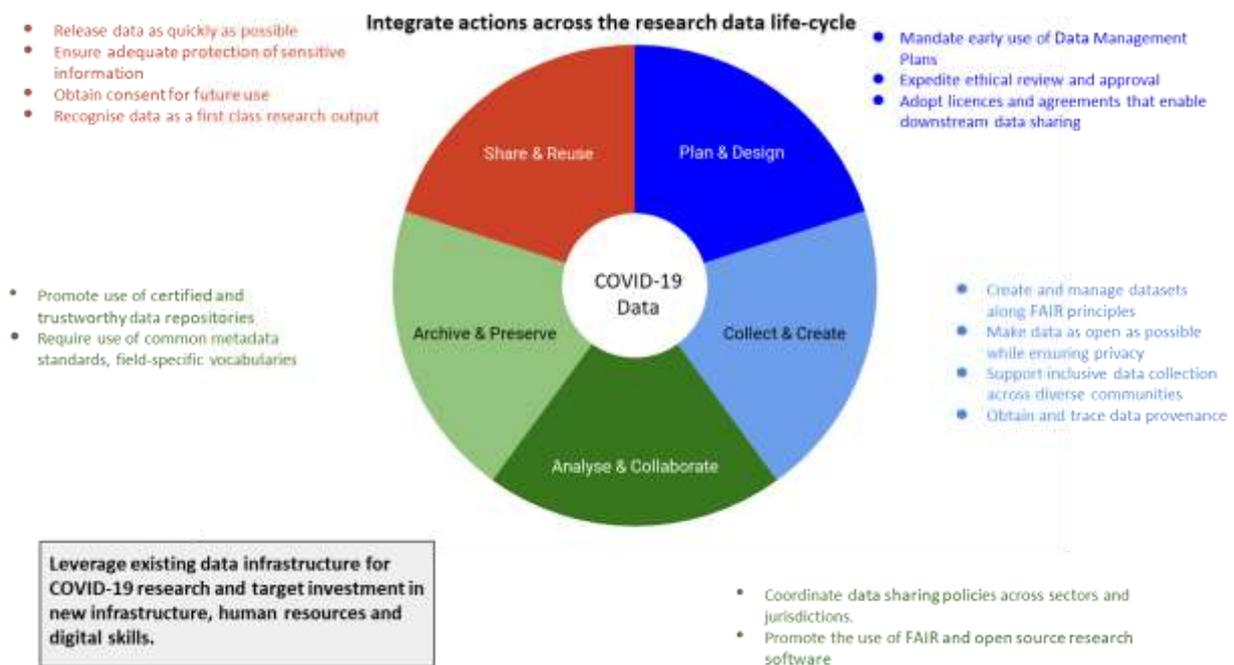
## International

A number of transnational frameworks have been established for the exchange of data and information during different types of crises (See OECD, 2018 [Scientific Advice During Crises](#)). These include the International Health Regulations (IHR), which are a transnational legal instrument that is binding on 196 countries across the World, including all WHO member countries. Under the IHR, countries are obliged to notify WHO of any event that may constitute a Public Health Emergency of International Concern (PHEIC). When the COVID-19 outbreak was declared as a PHEIC on January 30, 2020, the recommendations that were implemented under the IHR obliged all countries to share "full data'" with WHO and called on the global community to demonstrate solidarity and cooperation in this regard. To support this, the WHO provides a range of relevant guidance including its statement on [Developing Global Norms for Sharing Data and Results During Public Health Emergencies](#), the [Pandemic Influenza Preparedness Framework](#) and a [Code of conduct for open and timely sharing of pathogen genetic sequence data during outbreaks of infectious disease](#).

However, whilst the IHR and the various associated statements and guidance, provide a high level framework for international sharing of health-related data, their implementation depends on the political will, public health infrastructure and scientific capacity of individual countries. It also very much depends on having the right processes, standards and infrastructure in place to collect and manage the necessary data in such a way that it

can be trusted and made internationally comparable.  Ultimately, much of the responsibility for this falls on the scientific community and once the COVID-19 pandemic was was declared, a large number of national and international initiatives were rapidly developed to try and ensure that relevant authorities had the necessary data and analyses to respond to the pandemic (OECD, 2021).

Under the umbrella of the Research Data Alliance, a large international consortium of scientists, research managers and data specialists have been actively working together since the start of the pandemic to address the issues that are hindering data access across different scientific domains.  This work identifies a number of actions that are required throughout the research data lifecycle in order to enhance data access and re-usage (see fig 1).  Building on this, RDA has joined with a number of other international bodies, including CODATA and the World Data System, to develop the Virus Outbreak Data Network (VODAN), which focuses on making SARS-CoV-2 virus data FAIR and machine-readable.  The policy community, including governments, research funding agencies and institutional leaders, has a critical role to play to support such efforts and facilitate the rapid implementation of feasible solutions to enhance access to COVID-19 data for research.

**fig 1**



Integrate actions across the research data life-cycle

- Release data as quickly as possible
- Ensure adequate protection of sensitive information
- Obtain consent for future use
- Recognise data as a first class research output

- Mandate early use of Data Management Plans
- Expedite ethical review and approval
- Adopt licences and agreements that enable downstream data sharing

- Promote use of certified and trustworthy data repositories
- Require use of common metadata standards, field-specific vocabularies

- Create and manage datasets along FAIR principles
- Make data as open as possible while ensuring privacy
- Support inclusive data collection across diverse communities
- Obtain and trace data provenance

Leverage existing data infrastructure for COVID-19 research and target investment in new infrastructure, human resources and digital skills.

- Coordinate data sharing policies across sectors and jurisdictions.
- Promote the use of FAIR and open source research software

Many of the policy actions identified 'bottom up' by the international and interdisciplinary RDA community are targeted at research funders, a number  of whom had been working together internationally for several years prior to COVID-19 under the umbrella of the Global Research Collaboration for Infectious Disease Preparedness (GLOPID-R), to try and improve pandemic preparedness and data access.  GLOPID-R released its own Global roadmap for Data-sharing in Public-Health Emergencies in 2019.  These recommendations mainly focus on policy actions that funders can take to ensure that their grantees make their data available.  At the same time they recognise the need to work with other actors to "foster a culture and create and infrastructure where data sharing is an integral part of research."

It is reassuring that the generic policy actions that emerge from the RDA guidelines, which were developed in a bottom-up community-led initiative and specifically focus on COVID-19, are fully consistent with funders own guidelines and the 2021 OECD recommendation targeted at governments. One thing that has emerged very strongly from COVID-19 is the importance of science policy makers, funders and the scientific practitioner community working hand in hand and supporting each other to generate the knowledge and tools that are required to address the pandemic. This is absolutely essential when it comes to enhancing access to necessary data. Government mandates and funders guidelines are only useful in so far as they are adopted by the scientific community and translated into feasible practices that can be taken up widely.

## National

In order to be effective, International frameworks and guidance need to be embraced, and adapated by national authorities and institutions, who have the responsibility and resources and are in a position to establish mandates and incentives (sticks and carrots) to enable data access. Since the beginning of the COVID-19 pandemic, there have been a multitude of statements, declarations, and policy actions that align with those described in in fig 1 aimed at making COVID-19 related data FAIR. There have also been substantial new investments in data repositories, analytical services and cyber-infrastructure targeted at COVID-19 and in schemes to facilitate the use of these resources by researchers in both the public and private sector. For a summary of policy initiatives that were taken as part of the emergency response to the pandemic, up to June 2020, see Open science initiatives related to the COVID-19 pandemic.

# Domain-specific COVID-19 data sharing challenges

COVID-19 research crosses different domains of science, each of which have their own specific requirements and practices when it comes to the collection, management and use of data. Different domains have different 'data cultures' and vary considerably in the types of data that they handle and the extent to which they have adopted Open Science practices and exchange data internationally. This means that they are in different states of preparedness for making COVID-19 relevant data FAIR and so they may need proportionally more or less dedicated support and encouragement to achieve this.

### *Clinical medicine*

Timely sharing of data from clinical research, including clinical trials, is of utmost importance. Many clinical studies are being performed under enormous time pressure and, in the real world setting, experimental designs are not always fully optimal. At the same time there are compelling reasons to share even preliminary results that may have implications for treatment. This can only be justified when sufficient data is made available to allow verification by the wider scientific community. Ideally this would occur prior to final publication, e.g. using a pre-print process, but where this is not possible, **it is critical that the supporting data and analytical tools behind any claims that have implications for treatment are made openly available to scientific peers.** Many similar clinical trials for COVID-19 treatments are being planned or implemented in different countries. Registration of these studies in recognised international trial registries and the open sharing of data, and related documentation (e.g. protocols) is important to reduce duplication of effort, improve trial design and enable meta-analysis.

Clinical data are highly sensitive and their use is covered by legal frameworks and ethical requirements, including, most importantly, requirements for informed consent for both primary and secondary data use (see, for example, the Recommendation of the Council on health data governance , OECD 2016). There are many different types of

clinical information (personal and health data) and community standards exist to describe and structure most types and enable interoperability, although further work is urgently required in some areas. **Accredited data repositories, often federated in international networks, have well-established mechanisms for ensuring that clinical data can be preserved, appropriately documented, and reused in a secure, trustworthy and efficient manner and these repositories should be used for the deposition of clinical research data**.

### Omics

In order to understand the ways in which the SARS-CoV-2 virus causes the COVID-19 disease, research on biochemical processes at cellular and subcellular level is crucial. This includes research on the functioning of the virus as well as interaction with its host and encompasses genomics, proteomics and metabolomics.  Many research groups across with world are working in this area and a major determinant of the rate of progress is the speed at which data is shared. In this context, **it is imperative that omics data are preserved in domain-specific repositories that facilitate the consistent use of metadata and standards including field-specific vocabularies and ontologies**. In many cases these repositories also provide the data analysis tools and services that the community requires to extract knowledge from these data.

### Epidemiology

Understanding of the epidemiology of COVID-19 disease is crucial to inform public health policies to decrease infection rates and minimise deaths. Beyond the public health response, epidemiological data and models are critical for decision-making in relation to COVID-19 across all sectors of government.  Nevertheless, there is no international standard or coordinated system for collecting, documenting, and disseminating COVID-19 related data and metadata and, indeed, in some countries such data and the models that are used to analyse them, are not fully, openly accessible. The COVID-19 pandemic has highlighted the desirability of a data-driven, coordinated global system that encompasses preparedness, early detection, and rapid response to newly emergent infectious diseases.  Meanwhile, in the absence of such a complete system, it is critical that the basic data, which allows monitoring of the progress of the pandemic, including diagnostic testing data, is collected and made rapidly and openly accessible in all countries.  **Relevant data-holders should be supported and incentivised to coordinate their activities and establish common approaches across different jurisdictions**.

As vaccination progresses and we move from phase 3 clinical trials to phase 4 surveillance and reporting of adverse events it is critically important that the systems and standards are in place to allow rigorous and rapid integration of data from different countries and comparisons of different vaccines and vaccination protocols (Dean, 2021).  As new variants of the virus emerge, so they raise new questions about the effectiveness of vaccines (and other pharmaceutical and non-pharmaceutical interventions).  Timely and reliable answers will require international coordination and cooperation and the adoption of common standards and protocols. It will require the linkage of data from very large observational studies with omics data and clinical data, raising issues about technical and semantic interoperability, as well as careful attention to legal and ethical issues in an international context.

### Social sciences and humanities

Social scientists are collecting new information and reusing existing data sources, including on-line and social media, in research that can help inform policymakers about the social and economic impacts of COVID-19 and the measures that are being implemented to mitigate the effects of the pandemic.  Much of this research involves observational methods that produce unique data, which cannot be recreated in the future.   It is important that

data from social sciences research are collected, managed and preserved in ways that allow them to be reused across all domains and leveraged over the long-term. To achieve this, **researchers should ensure consent procedures, and adopt licences and agreements during data acquisition that enable downstream data sharing and preservation**.

Social media data present a particular challenge. These are a hugely valuable resource for exploring human behaviour, attitudes and responses to policy interventions but they are largely under the control of a small number of multinational companies, with substantial limitations on access and use by academic researchers. Where such data are available for research they raise particular challenges for obtaining informed consent and protection of privacy. A new OECD recommendation on Enhancing Access to And Sharing of Data will provide a high level policy framework to address some of these issues. For research using contact tracing apps, an open and transparent approach to data governance is required to build trust and promote adoption (see OECD policy note on [Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics](#))

The Covid-19 pandemic is not the first major global pandemic and will not be the last, the long-term data legacy from the COVID-19 pandemic should be an extremely valuable resource for future research in humanities and social sciences. Insights from earlier pandemics, have been extremely valuable in responding to COVID-19 and they could undoubtedly have been even more helpful if the relevant data had been systematically and sustainably managed over the long-term. However long term-data stewardship is not without cost and it will be important to identify what data is unique – cannot be collected again – and is most likely to be of use over the long-term. This is not straightforward as future usefulness cannot always be foreseen by those who generate or manage the data and so it is important that triage takes into account broad perspectives and likely advances in IT tools.

*Inter and trans-disciplinary research*

As the pandemic progresses, its cascading effects on all economic sectors and social activities have become clearer, whilst at the same time the complexity of these effects and the interactions between them mean that their immediate and longer term impacts have in many cases become more uncertain. Understanding this complexity requires the combination of expertise and data from many different scientific disciplines, including both natural and social sciences and humanities. It will also require expertise and insights from stakeholders outside the scientific community, including the public and private sector and citizens. Different disciplines and different stakeholders have different practices and expectations when it comes to data sharing and these will need to be reconciled. The benefit to society and global public good perspective needs to be balanced against legitimate data protection concerns, with the aim of making FAIR data as 'open as possible and as closed as necessary'.

[The RDA-COVID-19 WG Recommendations and Guidelines for Data Sharing](#) provide detailed advice for different research domains. This includes guidance on data and metadata standards, controlled vocabularies, trustworthy repositories, data licensing and data documentation. It also provides concrete guidance on participant consent and protocols for managing personal data, as well as advice on overarching legal and ethical considerations.

# Areas for policy action [to be further developed after the workshop]

Building on the work by OECD and RDA, 4 key areas can be identified where policy-makers need to act now to enable access to data, and associated models, software and code for research on COVID-19 and improve preparedness for future pandemics. Effective actions can also mark a significant step in enhancing access to research data as part of the overall transition to Open Science.

## 1.    Provide leadership and support for coordinated, cross-jurisdictional efforts to foster data sharing:

Immediate actions (COVID-19)

- Policymakers, funders and institutions urgently need to update data sharing policies and processes across all domains in government, healthcare systems, and research institutions to support FAIR data access for COVID-19 research and analysis.
- Policymakers and research funders should actively engage with digital technology companies, mobile network operators, social network companies and others in the private sector who hold data that can better help understand the pandemic and population behaviour. Data sharing policies should be adopted, based on existing best practices, to encourage and facilitate data flows between data holders and the research community, whilst ensuring the protection of citizens' rights.

Future perspective

There is a need to update policies and processes for cross-sectoral access to data for research in preparation for a range of future crises.  Whilst some access provisions may only be applicable in emergency situations, arrangements that are tested and prove effective now should be adopted routinely in the future unless there are good reasons not to do so.

## 2.    Maximise the use of existing investments in digital research infrastructure and skills:

Immediate actions (COVID-19)

- Existing data management infrastructure should be leveraged to support COVID-19 research and used as a basis for building new capacity, as required. Economies of scale and federated models should be considered when planning institutional, disciplinary, sector-wide, or regional/national data infrastructure to limit duplication, encourage collaboration, and maximise return on investment.
- Policymakers, funders and institutions should invest in the human resources required to maintain digital infrastructure and FAIR data provision for research on COVID-19 and future pandemics, and support dedicated user-training programmes.

Future perspective

Investment in infrastructure, digital skills and resources for data management should be made strategically so that all relevant jurisdictions and sectors are equipped to make significant contributions towards the evidence base for pandemic response and are better prepared for future crises. This is a baseline requirement for Open Science and the current pandemic is shedding light on gaps and weaknesses that need to be addressed at multiple scales.

## 3.    Make data available in a FAIR, responsible and timely manner:

Immediate actions (COVID-19)

- Funding mechanisms for COVID-19 should mandate the development of data management plans and include support for data stewardship to ensure that data are made FAIR.
- Policymakers and funders should promote and/or mandate the use of trustworthy data repositories that have been certified, are subject to rigorous governance, and committed to longer-term preservation of their data holdings. This is particularly important with regard to sensitive and personal data;
- In the context of COVID-19, curated datasets should be recognised as first-class research outputs equal in value to traditional peer-reviewed articles.

Future perspective

The Covid-19 pandemic can act as a catalyst to accelerate the implementation of these actions that recognise the value of FAIR data and of the professionals that are required to generate this data.

## 4.     Implement robust and transparent data governance mechanisms that ensure rigorous attention to ethical and legal considerations and promote trust with data providers, data users and the public at large

Immediate actions (COVID-19)

- Policies and mechanisms for research data governance should be designed to ensure that sensitive data are handled in a trusted and secure way, that provenance can be traced and that data from different populations are collected and processed in ways that are representative and ethically appropriate.

- Established legal and ethical principles for research, in particular requirements for informed consent and respect for privacy, should be enforced during the pandemic. Any derogations from normally applicable principles, which might exceptionally be justified on the basis of public interests, should be transparent and time limited. See OECD policy note on Ensuring Data Privacy as we battle Covid-19

- Research on COVID-19 should be inclusive of different communities, which has implications for study design, data collection and stewardship. Policymakers should adopt the CARE Principles for Indigenous Data Governance, which set minimum standards for collectors, users and stewards of data and point to the need for Indigenous Peoples and nations to be engaged in governance on their own terms across COVID-19 data lifecycles and ecosystems.

- Funders and institutions should make sure that measures to manage risk (anonymization, aggregation, data-use agreements, safe havens) are used to make access as simple as possible, while providing adequate protection of sensitive information.  Ethics committees and institutional review boards, should be fully supported and have ready access to the necessary legal and digital skills to fully assess data-related risks and advise on mitigation strategies.

Future perspective

The COVID-19 pandemic has raised new challenges for data governance and surfaced existing problems that require urgent attention.  Experimentation is taking place in this regard– established procedures are being scaled up or adapted to new requirements and new mechanisms and agreements are being established.  It will be important that best practices are identified and shared and become embedded in future data governance processes.

## Concluding remarks

Access to data has been essential for the scientific community to respond effectively to the COVID-19 pandemic and this will surely be the case for future crises.  The pandemic has come at a time when a transition towards Open Science was gaining momentum and many international frameworks and national policy measures were already  in place with the aim of enhancing access to research data and associated tools and services. However, this transition was moving at different speeds in different parts of the research community and in different countries. The pandemic has demonstrated the need to accelerate the process across all research domains and national jurisdictions.  Policy-makers have taken up this challenge and implemented many targeted measures to promote access to COVID-19 data.  The research community has also responded at an unprecedented scale and more research data is being effectively shared within and across different domains and sectors than has ever been the case before.  Solutions have been found to many of the complex technical, legal and ethical challenges that

have arisen. There has been recognition of the importance of sustainable data infrastructure and the personnel who are required to manage this infrastructure, as well as the skills that are required to perform data-intensive science.

The pandemic is not over and, as it evolves, it generates new research needs and new requirements for data sharing across different domains and countries. This is both a major challenge and an opportunity. It is an opportunity to properly implement measures that have already been agreed internationally to build the global research data infrastructure, processes and agreements that are essential in times of crises and beneficial for science and society in normal times.