Enhancing access to research data to combat COVID-19

Recommendations to Funders

Background

In March 2020, the European Commission made a direct request to RDA to create global guidelines and recommendations for data sharing under COVID-19 circumstances. A working group was created, over 600 data professionals and domain experts signed up, and work began in early April 2020. Subgroups were created for four research areas (clinical data, omics practices, epidemiology and social sciences) and four cross-cutting areas (legal and ethical considerations, research software, community participation and Indigenous data). The resulting work, *RDA-COVID-19 WG Recommendations and Guidelines for Data Sharing*, published in June 2020, is a rich set of detailed guidelines to help researchers and data stewards follow best practices, to maximise the efficiency of their work, and to act as a blueprint for future emergencies.

The publication provides detailed advice 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. The detailed researcher guidelines are complemented by higher level recommendations to help policymakers and funders maximise timely, quality data sharing and to formulate appropriate responses during public health emergencies. These recommendations are the focus of the briefing below, and are meant to provide expert advice to funders on how to support and foster the best possible approach to research dissemination at a systems level, while also providing guidance on shaping grant conditions for COVID-19 research that will promote timely and reusable sharing of research outputs.

Introduction

Timely 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 code and software that is required for the production and analysis of this data, is necessary across the many different domains of science that are being mobilised to understand and effectively combat COVID-19. Immediate results of research are informing public health strategies, and data and analytical tools are also essential for designing, monitoring and assessing the impact of the socio-economic policies being implemented in response to the pandemic.

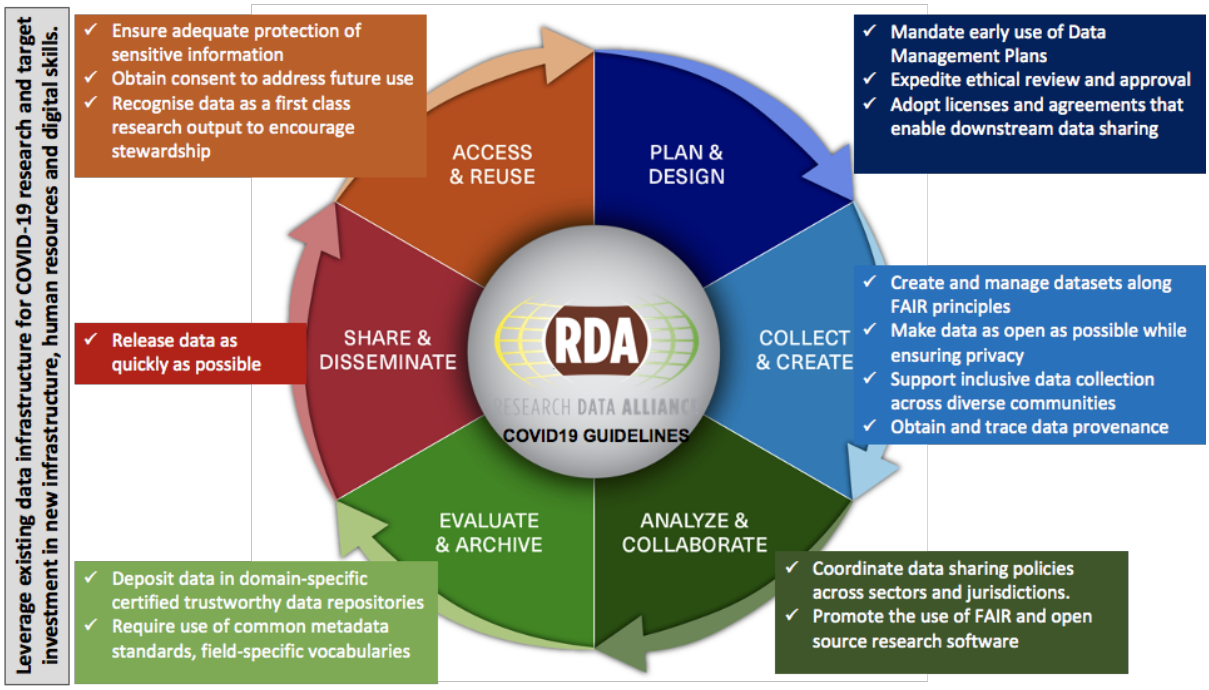
The policy community, including governments, research funding agencies and institutional leaders, has a critical role to play to support sharing efforts and facilitate the rapid implementation of feasible solutions. In the short-term this will mainly require mandates and incentives. In the medium to longer-term there are also issues related to training and skills and investment in infrastructure.

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 these diverse systems is currently a major obstacle to timely and effective access. In short, a lot of data are not yet, or not sufficiently, findable, accessible, interoperable and reusable (FAIR). The unprecedented spread of the virus has prompted a rapid and massive research response. However, 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 solutions exist.

To make the most of global research efforts, scientific findings and data need to be shared rapidly, in a way that is useful and comprehensible. The FAIR and timely sharing of data and associated software is an essential element of the Open Science approach that the world needs to effectively combat pandemics like COVID-19. Limiting access not only slows up scientific progress but, as we have seen recently, can undermine public trust in science and science-based decision-making. Policymakers, research funders, and research institutions around the world must work together, to mandate and support policy actions that harmonise and streamline the timely provision and exchange of data locally, nationally and internationally.

Plotting Policy Recommendations of the Research Data Lifecycle

This infographic notes how policy actions are necessary across the different steps of the research data lifecycle to ensure enhanced access to Covid-19 research data



Source: Inspired by the Harvard Biomedical Research Data Lifecycle

Key recommendations to Funders

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

- 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 and open 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.

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

    - 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 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.

    - 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.

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

    - FAIR data provision and timely release of data should be a conditional requirement for all publicly-funded COVID-19 research.

    - 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.

    - Funders and institutions should provide support and recognition for data stewardship and digital research support professionals within COVID-19 projects.

    - Where sensitive (e.g. personal) data is involved and research proposals are considered by institutional review boards and/or ethical committees, these structures should have the necessary data management expertise.

    - 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;

    - Policymakers and funders should implement a “data first” publication policy by encouraging the publication of COVID-19 data articles in “open” peer-reviewed data journals and/or mandating and supporting the deposit of data and associated code in certified trustworthy data repositories in tandem with the publication of articles;

    - In the context of COVID-19, curated datasets should be recognised as first-class research outputs equal in value to traditional peer-reviewed articles.

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

- 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.
- During the pandemic, ethical review and approval for using and sharing research data should be expedited, and, regardless of the societal interests, measures to protect privacy of individuals should be paramount. See also OECD policy note on *Ensuring Data Privacy as we battle Covid-19*.
- For research using contact tracing apps, an open and transparent approach to data governance is required to build trust and promote adoption. See also OECD policy note on *Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics*
- 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 (anonymisation, aggregation, data-use agreements, safe havens) are used to make access as simple as possible, while providing adequate protection of sensitive information.

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 extent to which they have adopted Open Science practices. 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 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 and improve trial design.

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. 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 to understand 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 the 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.

Epidemiology

Understanding of the epidemiology 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 emerging 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 to coordinate their activities and establish common approaches across different jurisdictions.

Social sciences

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.

Contact

Natalie Harrower, | Director, Digital Repository of Ireland | n.harrower@ria.ie
Ingrid Dillo, RDA COVID-19 WG CoChair | Deputy Director, DANS | ingrid.dillo@dans.knaw.nl
(Co-Chairs of the RDA COVID-19 working group)

Further Reading

Full Recommendations and Guidelines: https://doi.org/10.15497/rda00052
The Value of RDA for COVID-19: https://www.rd-alliance.org/value-rda-covid-19-0