RDA COVID-19 Working Group

Recommendations and Guidelines

1st release

24 April 2020

RDA Recommendation (1st Release - 24 April 2020)
Produced by: RDA COVID-19 Working Groups, **2020**

https://rd-alliance.org/groups/rda-covid19

Document Metadata

Identifier	DOI: 10.15497/RDA00046
Title	RDA COVID-19; recommendations and guidelines, 1st release 24 April 2020
Description	This is the first draft of the "overarching" RDA COVID-19 Guidelines document, and is intended to provide an update on the progress of the WG, as well as a focus on high-level recommendations that run across all 5 sub-groups in this initial effort, as well as the cross-cutting theme of Research Software.
Date Issued	2020-04-24
Version	Draft guidelines and recommendations; first release, 24 April 2020, version for public review
Contributors	RDA COVID-19 Working Groups
Licence	

History of the discussions in the Working Groups that led to this document can be viewed in the comments made in the associated Google documents.

This work was developed as part of the Research Data Alliance (RDA) 'WG' entitled entitled 'RDA-COVID19,' 'RDA-COVID19-Clinical,' RDA-COVID19-Community-participation,' 'RDA-COVID19-Epidemiology,' 'RDA-COVID19-Omics,' 'RDA-COVID19-Social-Sciences,' and we acknowledge the support provided by the RDA community and structures, and contributions by the Research Software Alliance <ReSA>.

Table of Contents

1. Function of the RDA COVID-19 WG

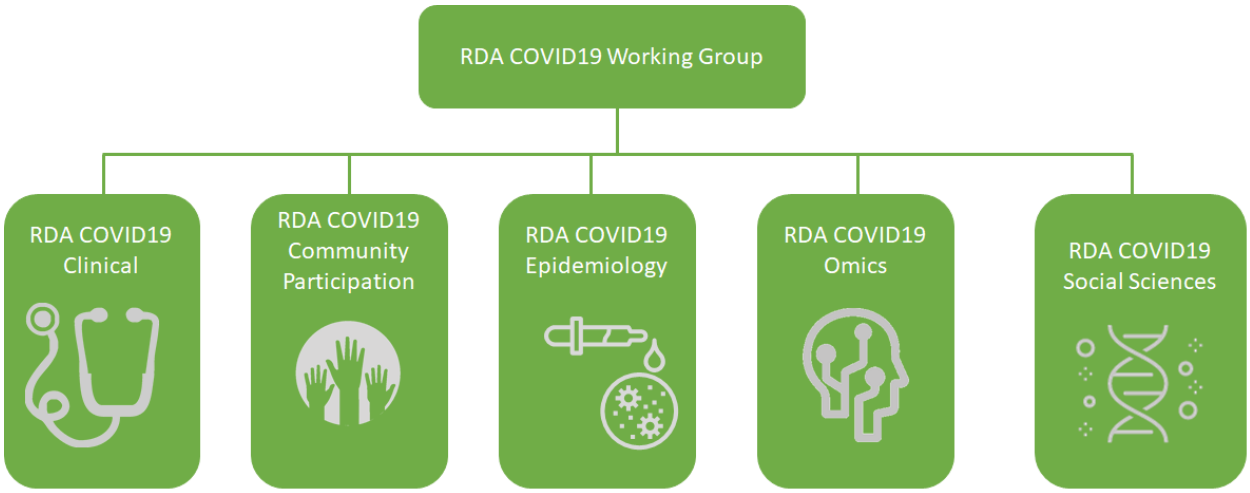
During a pandemic, data combined with the right context and meaning turns into knowledge for informing public health response. The insights derived from the knowledge are used by policy makers to make informed decisions for rapid response. Timely collection, reporting and sharing of data with the research community, public health practitioners, clinicians and policy makers will inform assessment of the likely impact of a pandemic to implement efficient and effective response strategies.

Public health emergencies clearly demonstrate the challenges associated with rapid collection, sharing and dissemination of data and research findings to inform response. There is global capacity to implement systems to share data during a pandemic, yet the timeliness of accessing data and harmonisation across information systems are currently major roadblocks. The World Health Organisation's (WHO) statement on data sharing during public health emergencies clearly summarises the need for timely sharing of preliminary results and research data. There is also a strong support for recognising open research data as a key component of pandemic preparedness and response.

The objectives of the RDA COVID-19 Working Group (CWG) are:

1. to clearly define detailed guidelines on data sharing under the present COVID-19 circumstances to help stakeholders follow best practices to maximize the efficiency of their work, and to act as a blueprint for future emergencies;
2. to develop guidelines for policymakers to maximise timely data sharing and appropriate responses in such health emergencies;
3. to address the interests of researchers, policy makers, funders, publishers, and providers of data sharing infrastructures.

The CWG is addressing the development of such detailed guidelines on the deposit of different data sources in any common data hub or platform. The guidelines aim at developing a system for data sharing in public health emergencies that supports scientific research and policy making, including an overarching framework, common tools and processes, and principles that can be embedded in research practice. The guidelines to be developed will address general aspects related to the principles that data should adhere to, for example FAIR and other principles. The initial work was divided into 5 thematic areas as a way to focus the conversations, and provide an initial set of guidelines in a tight timeframe. Additional themes and details will be added over time in iterative releases of this document.



2. Status of the RDA COVID-19 WG Effort

The RDA COVID-19 WG was initiated after a conversation between the RDA Secretary General and European Commission contacts. The first meeting to determine the work package was held on March 20th, and included a number of RDA stakeholders. Subsequent to this, the Secretary General reached out to colleagues in the RDA community to act as Co-Chairs, and the first meeting of this group was held on March 30th. The next step was to invite a group of Moderators to facilitate the discussion of the 5 sub-groups, and the first group meetings started taking place soon after.

As of mid-April, there were well over 400 members of the CWG, relatively evenly spread across the 5 groups. This preliminary April 24th draft represents the earliest set of draft outputs from the first set of sprints, and should be understood as an early indication of progress in the space of 3 weeks, with more details and final outputs to follow. The various sub-group drafts collected here reflect different approaches and initial work efforts: subsequent drafts will synchronize these efforts into a more integrated series of guidelines. These 5 thematic guidelines will be enhanced with a series of "cross-cutting" guidelines, that articulate principles and recommendations common across all 5 areas. One initial cross cutting guideline is included here: Research Software. Additional cross-cutting guidelines will include Legal and Ethical - suggestions for others are welcome.

In the spirit of the RDA community and its open process, we are seeking feedback from the COVID-19 WG members, as well as the broader community, early and often during this process. This feedback will inform our work and will be incorporated into the sub-group discussions, and the next set of writing sprints.

This Working Group and the subgroups operate according to the RDA guiding principles of Openness, Consensus, Balance, Harmonization, Community-driven, Non-profit and technology-neutral and are OPEN TO ALL.

3. Foundational Principles/Recommendations

The 5 thematic sub-groups have each produced a first iteration of the challenges facing researchers working on COVID-19, as well as recommendations/guidelines for improving data sharing; these subgroup guidelines should be considered directly depending on the relevant area of COVID-19 research. However, certain foundational aspects appear across these subgroups, so we present these here as foundational elements that apply across all themes.

3.1 Challenges

3.1.1 Rapid Pace of Research Under the Pandemic

The unprecedented spread of the virus has prompted a rapid and massive research response, but to make the most of global research efforts, findings and data need to be shared equally rapidly, in a way that is useful and comprehensible. New findings and understandings need to be disseminated and built upon at a pace that is faster than usual, because decisions are being taken by healthcare practitioners and governments on a daily basis, and it is urgent that they are well-informed.

Emerging infections are largely unpredictable in nature and there is limited data to support disease investigation. The evidence base generated from early outbreak data is critical to inform rapid response during an emerging pandemic. Lack of pre-approved data sharing agreements and archaic information systems hinder rapid detection of emerging threats and development of an evidence-based response.

3.1.2 Critical Need for Data Sharing

The COVID-19 pandemic has revealed how interconnected we are globally, and how interdependent we are in terms of research, public health, and economy. Data in relation to this pandemic is being collected and created at a high velocity, and it is critical that we can share this data across cultural, sectorial, jurisdictional, and disciplinary boundaries.

3.1.3 Lack of Coordinated Standards and Context

While the research and data are abundant, multi-faceted, and globally produced, there is no universally adopted system, or standard, for collecting, documenting, and disseminating COVID-19 research outputs, and many outputs are not reusable by, or useful to different communities, if they have not been sufficiently documented and contextualised, or appropriately licensed. There is an urgent need for data to be shared with minimal contextual information and harmonised metadata so that it can be reused and built upon (see the OECD Open Science Policy Brief).

3.2 Recommendations

3.2.1 FAIR and Timely

The consensus in this series of guidelines is that research outputs should align with the FAIR principles, meaning that data, software, models and other outputs should be Findable, Accessible, Interoperable and Reusable. However, there is also consensus that outputs need to be shared as quickly as possible in order to have a direct impact on the progress of the pandemic. A balance between achieving 'perfectly' FAIR outputs and timely sharing is necessary with the key goal of immediate and open sharing as a driver. Researchers should be paired with data stewards to facilitate FAIR sharing, and data management should be considered at the start of a study or trial. Immediate open access with open licenses is desirable, but some effort should be put into the quality and documentation of the dataset.

3.2.2 Metadata

Data must be accompanied by openly accessible metadata so that it can be discovered, interpreted correctly, and reused for subsequent research. While rich metadata is desirable, even a minimum set of key fields/descriptors is valuable. The use of common metadata standards, as adopted by one's relevant discipline, as well as vocabularies, are highly recommended, and metadata should describe the data as well as the terms under which it can be accessed and reused. Ideally, data and metadata should be exposed via machine readable endpoints (e.g. RDF, APIs) to facilitate analysis and research on the data, especially for machine learning and other statistical methods. Where there are restrictions on accessing or using datasets, metadata should be shared openly to enable discovery (e.g. CC0 or CC-BY licenses).

3.2.3 Documentation

Research outputs need to be documented, which includes documentation of methodologies used to define and construct data, data cleaning, data imputation, data provenance and so on. Software should provide documentation that describes at least the libraries, algorithms, assumptions and parameters used. Equally, research context, methods used to collect data, and quality-assurance steps taken are important. When sharing datasets, other relevant outputs (or documents) should also be made available, such as codebooks, lab journals, or informed consent form templates, so that data can be understood and potentially linked with other data sources.

3.2.4 Use of Trustworthy Repositories

To facilitate data quality control, timely sharing and sustained access, data should be deposited in data repositories. Whenever possible, these should be trustworthy data repositories (TDRs) that have been certified, subject to rigorous governance, and committed to longer-term preservation of their data holdings. As the first choice, widely used disciplinary repositories are recommended for maximum accessibility and assessability of the data, followed by general or

institutional repositories. Using existing open repositories is better than starting new resources. By providing persistent identifiers, demanding preferred formats, rich metadata, etc., certified trustworthy repositories already guarantee a baseline FAIRness of and sustained access to the data, as well as citation.

3.2.5 Ethics & Privacy

The ethical and privacy considerations around participant and patient data are significant in this crisis, and several guidelines note the need to find a balance that takes into account individual, community and societal interests and benefits whilst addressing public health concerns and objectives. Access to individual participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse.

3.2.6 Legal

Technical solutions that ensure anonymisation, encryption, privacy protection, and data de-identification will increase trust in data sharing. The implementation of legal frameworks that promote sharing of surveillance data across jurisdictions and sectors would be a key strategy to address legal challenges. Emergency data legislation activated during a pandemic needs to clearly outline data custodianship/ownership, publication rights and arrangements, consent models, and permissions around sharing data and exemptions.

4. Clinical Sub-Group Guidelines

4.1 Sub-Group Focus and Description

Clinical activities are at the forefront to combat the COVID-19 pandemic. Although many aspects of such activities were considered in the scope of the sub-group, the work of the Clinical Subgroup centers first on how clinical trials are conducted and how clinical information and results are shared and consumed in a trustworthy and efficient manner.

4.2 Initial Sub-Group Guidelines

4.2.1 Clinical trials on COVID-19

Clinical trials are an important research area to discover and make available safe and effective treatments for COVID-19. International, regional, and national legal and methodological frameworks exist for clinical trials, that also take into account ethical principles. Specific recommendations on registering, performing, and sharing ongoing clinical research are the following:

1. Clinical trials in COVID-19 should be registered at or before the time of first patient enrollment and protocols possibly published in order to favor

harmonization of studies, collaboration among centres as well as to avoid duplication of effort.

2. Multi-centre/multi-country studies including a high number of patients should be recommended to generate sound evidence on COVID-19 treatments. Collaborative trials and multi-arm studies comparing different drugs are advisable.
3. Lawful fast track approval procedures of clinical trials in cases of public health emergencies exist that speed up processes while protecting individual rights. Platforms that point to them in the various national and international institutions should be further developed and administrations should apply them diligently and transparently.
4. Heterogeneity of registries regarding the number of studies listed and the information available for individual studies should be overcome through a dialogue among different platforms.
5. Protocols should follow standard criteria for data collection, stratification of the randomized population, type of intervention and comparator, a minimal set of primary outcome measures (e.g. SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials).
6. Availability for timely publication of results - even for negative and withdrawn studies - and for data sharing should be declared by investigators and sponsors at the time of study registration and included in the study documents (e.g. protocol, patient information and consent form).
7. Regulatory bodies should facilitate compassionate use of approved repurposed drugs and establish a fast track for approval of all COVID-19 drugs. Adaptive study designs and post-authorization efficacy and safety studies after exceptional or conditional approval should be planned with sponsors in order to favor early access of severe patients to promising medicines.
8. Adequate tools should be implemented for collection and analysis of reliable real-world data of drugs approved for the treatment of COVID-19.
9. Data and trial documents should be made available for sharing. Individual participant data sharing should be based on explicit broad consent by trial participants (or if applicable by their legal representatives) to the sharing and reuse of their data for scientific purposes, according to applicable law. Where real-world data are collected from patient registries or similar data sources not involving specific consent to participate, patients' privacy must be adequately protected. Data and trial documents should be transferred to a suitable and secure data repository to help to ensure that the data are properly prepared, are available in the longer term, are stored securely and are subject to rigorous governance. Repositories that explicitly support data sharing for COVID-19 trials should be announced (e.g. Vivli).
10. For data sharing, clinical trial data should always be associated with adequate and standardised metadata to improve discoverability ("F" in FAIR).

11. In order to speed up the process of data sharing, standardised agreements for sharing of data from COVID-19 trial should be developed and implemented (e.g. data transfer agreements, data use agreements).
12. Tools should be developed to enable regular harvesting of metadata objects from clinical trials, allowing identification of trials and all related data objects (e.g. protocol, data set, a summary of results, publication, data management plan) through one portal.

More information is at clinical trials and clinical aspects documentation.

4.2.2 Trustworthy sources of clinical data

During a pandemic like COVID-19, it is critical to spend limited time and resources on reliable data sources that provide data and metadata of high quality and guarantee the authenticity and integrity of the information. The recommendations are:

1. Trustworthy repositories should be leveraged as a vital resource for providing access to and supporting the depositing of research data. However, as an emerging and evolving area in biomedical domains, trustworthiness assessment should not be limited to certification or accreditation. A wide-range of community-based standardized quality criteria and best practices should also be considered.

Additional materials from the Working Group can be found at: RDA-COVID19-Clinical.

4.3 Additional Working Documents & Links

- Clinicaltrials.gov
- https://www.nlm.nih.gov/NIHbmic/index.html
- https://datascience.nih.gov/covid-19-open-access-resources
- Sharing and reuse of individual participant data from clinical trials: principles and recommendations
- CDISC interim User Guide for COVID-19

5. Community Participation Sub-Group Guidelines

5.1 Sub-Group Focus and Description

The context in which we work — data and community participation
Public health emergencies require profound and swift action at scale with limited resources, often on the basis of incomplete information and frequently under rapidly evolving circumstances. The current COVID-19 pandemic is one such emergency, and its scale is unprecedented in living history. Worldwide, many communities are coming together to address the emergency in a plethora of

ways, many of which involve data in various fashions. For instance, they produce or mobilize data, add or refine metadata, assess data quality, merge, curate, preserve and combine datasets, analyze, visualize and use the data to develop maps, automated tools and dashboards, implement good practices, share workflows, or simply engage in a range of other activities that can or do leave data traces that can be leveraged by others.

While emergency-triggered sharing goes back millennia, data sharing is a relatively new aspect of emergency response, and the size, scale and complexity of the data relevant to the current pandemic are many orders of magnitude greater than even those of other recent epidemics, e.g. SARS, MERS, Zika or Ebola. This abundance of data, while in our favour in principle, can also be our Achilles heel if we - and our technology - are not able to openly share, understand and combine this data to gain the maximum insights it can provide, and to communicate those insights to the communities for which they are relevant and to the wider public.

The aims of the RDA-COVID-19 Community Participation subgroup

Our primary aim is to support the work of communities which are sharing data with the goal of improving research outputs and public knowledge. To achieve this, our objectives include highlighting the achievements and outputs of groups who practice sharing and to broaden access to the existing guidelines for sharing best practices. As described in "Principles of data sharing in public health emergencies" and similar publications, guidelines address issues of giving credit for contributions, legality in sharing data, technical considerations in making data Findable, Accessible, Interoperable and Reusable (FAIR), or other similar guidance for collaborating in research during a crisis.

With this objective in mind, the subgroup seeks to also take on an active role of bridging communities and ensuring inputs are streamlined, perspectives from communities are considered, and the collaborative outputs of all the RDA COVID-19 subgroups are widely communicated. The aim of linking communities and supporting communication is also designed to help coordination and avoid duplication of efforts since many communities are driving similar or complementary efforts to help the response to the current public health emergency.

These guidelines aim to facilitate the timely sharing of data relevant to the COVID-19 response and build much-needed capacity for similar events in the future. An effective and efficient response to a public health emergency, such as the current pandemic, demands and holds immense value for both public and science communication, informing opinions and understanding, whilst supporting decision-making processes.

Although these principles have been developed with research data in mind, it is also desirable that data created directly by citizens (be that in a role as citizen

scientists or not), patients, communities and other actors in a health emergency be produced, curated and shared in line with the spirit of these sharing principles. For example, community projects such as OpenStreetMap and Wikidata generate very valuable FAIR and open data, which can be analysed and used along with data from professional research and other sources.

5.2 Initial Sub-Group Guidelines

5.2.1 Stakeholders

The intended audience for this subgroup's outputs includes

1. ***Researchers*** undertaking activities along the entire life-cycle of pertinent data, especially those not covered by the other RDA COVID-19 WG subgroups and involving broad-scale community participation but also data stewardship of the community-generated data.
    1.1. **Citizen scientists** undertaking research activities and in need of guidance (e.g. in terms of ethics) as well as means to seamlessly contribute to a common body of knowledge and collaborate with other actors involved.
2. ***Policymakers*** are involved in setting the framework for community participation, funding innovation, working on research policy or focusing on integrating data in decision making.
3. ***Patients***, caregivers and the communities around them that are involved in leveraging data to improve prevention, diagnostics or treatment (this complements the work of the RDA COVID-19 Clinical subgroup).
4. ***Developers*** involved in the creation or maintenance of applications targeted at community data collection that are specific to COVID-19 (e.g. contact tracing apps or exposure risk indicator apps) or more generic in nature (e.g. health or neighbourhood apps).
5. ***Device makers*** involved in developing sensors and data generating products for the community to use.
6. ***Communicators*** involved in informing communities and societies at large about data-related aspects of the COVID-19 pandemic, translating data into meaningful and easy to grasp information, and circulating graphics or key messages in conventional or social media.
7. ***Citizens and the public at large***, i.e. members of any community wanting to contribute to the COVID-19 response in ways that involve data and who want to have a say in how to balance that with legal and ethical issues surrounding such data.
8. ***Other actors (individuals or organisations)*** who are involved in community-based activities around COVID-19 related data.

5.2.2 Our approach going forward

Whenever possible, we aim to reuse and share applicable recommendations that already exist for specific communities and/or types of data. To this end, we will adopt a standardised approach to identify existing guidance related to specific use cases in communication with relevant communities.

For existing guidance, the subgroup aims to collaborate with relevant communities to review and help refine it and support a broader distribution. If guidance is needed but not available yet, the subgroup will help identify issues and support drafting applicable recommendations. Beyond that, we encourage community members to help translate such recommendations (i) between languages; (ii) from prose into practice, including code and other formalized workflows; (iii) from one community or data type to similar ones.

Topics that we anticipate to be relevant in the context of the above-mentioned use cases include but are not limited to: collaborative data collection, collaborative service or software development initiatives, crowdsourcing of data curation services, data sovereignty when sharing across communities, citizen-led community responses, participatory disaster response strategies, digital platforms or apps to enable public participation and/or offer open data, digital tools to enable public participation.

Furthermore, the group plans to leverage the strengths of the RDA as an international community of data specialists and practitioners as well as reach out beyond to ensure expert input in addressing overarching topics such as ethics and social aspects, indigenous data, global open research commons, metadata standards, persistent identifiers and scientific annotation.

5.3 Additional Working Documents & Links

- RDA-COVID-19
- RDA-COVID-19 Community Participation
- Initial scoping doc for Community participation recommendations
- Parent document of RDA COVID-19 WG
- Root folder

Additional materials from the Working Group can be found at: RDA-COVID19-Community-participation

6. Epidemiology Sub-Group Guidelines

6.1 Sub-Group Focus and Description

Responses to the COVID-19 pandemic have been massive and multifaceted worldwide. An immediate understanding of the disease's epidemiology is key to slowing infections, minimizing deaths, and making informed decisions about when, and to what extent to impose quarantine measures, and when and how to reopen society. As economies "open up," improved and innovative surveillance and follow-up will be key to minimizing resurgence.

There is no standard or coordinated system for collecting, documenting, and disseminating COVID-19 related data and metadata, making their reuse for timely epidemiological analysis challenging due to issues with documentation, interoperability, completeness, and reliability of the data.

The key elements that block sharing and reuse of epidemiology data are common across many domains. These include non-machine-readable data (e.g., PDF), heterogeneous measurement standards, divergent metadata formats, lack of version control, fragmented datasets, delays in releasing data, non-standard definitions and reporting parameters, lack of metadata, unavailable or undocumented computer code, frequently changing web addresses, copyright and usage conditions, translation requirements, consents, approvals, and legal restrictions. In addition, clinical, eHealth, surveillance, and research systems within and across jurisdictions or providers do not integrate well, or at all.

A major difficulty at this time is the lack of contextual data needed to study the evolution of disease in sub-populations. They include, among others, otherwise healthy sub-populations that are vulnerable to serious long term effects following recovery that we don't know about yet because we don't have the data and because we are focusing on deaths. They also include age-specific vulnerabilities, disadvantaged sub-populations with limited health care, vulnerabilities evident in severe disease associated with comorbidities, and vulnerabilities due to environmental conditions, and due to social and cultural norms. Vigilance will be necessary to follow sequelae and immunity. These data are not collected systematically in the healthcare system or via different survey instruments. Moreover, merging clinical databases with other types of databases is difficult or impossible due to interoperability and legal reasons.

The present crisis demonstrates more than ever before just how intimately connected and interdependent the world is across countries and organizations. It also lays bare the stark reality and shortcomings of our largely antiquated data systems and data sharing agreements within and between domains that severely hinder rapid detection of emerging threats and development of a science-based response to them. Barriers are encountered between countries and between jurisdictions within countries, and between national and international organizations. The epidemiology of COVID-19 is dependent on input data from across a wide variety of domains that include not only clinical and surveillance data, but also administrative, demographic, socioeconomic, and environmental data amongst others. A process of scientific data modernization and related policies in all of these domains is urgently needed to support epidemiologic analyses and modeling that provide critically important insight and understanding of the newly emergent SARS-2 virus and COVID-19 disease that it causes.

Implementation of the principles and tools of Open Data and Open Science (e.g. Open Access and FAIR data) that have been under development for several years would solve many of these problems. While science has been gradually moving in this direction, it will require a concerted effort by governments, policy makers, research institutions, clinicians and scientists worldwide to achieve the culture change needed for full adoption. The COVID-19 pandemic highlights the urgent need to remove barriers and accelerate this process now to better respond to the current need for rapid discovery, acquisition and integration of relevant data, and sharing of accurate data to support evidence-informed public health decisions during this rapidly evolving catastrophe.

6.2 Initial Sub-Group Guidelines

1. Internationally harmonized COVID-19 policies based on empirical modeling and epidemiological evidence.
2. Account for public health decision making in modelling COVID-19 inputs and outputs.
3. Rapid development of a consensus standard for minimally viable COVID-19 surveillance data:
    a. Definition of and criteria for COVID-19 testing, and reporting on testing.
    b. Policies and definitions: interventions, contact tracing, reporting of cases, deaths, hospitalizations and length of stay, ICU admissions, recoveries, reinfections, time from detection to death or recovery, comorbidities, followup to identify serious long-term effects in recovered cases, location, demographic, socioeconomic information, and outcome of resolved cases.
    c. Daily reporting cutoff times.
4. Harmonized approaches to comparably quantify side-effects of pandemic mitigation measures on our societies such as shifts in morbidity, mortality, health care utilisation, quality of life, social isolation.
5. Rapid development of an internationally harmonized specification to enable the export/import of epidemiologic data from clinical systems, record linkage to population-based surveillance data, and automatic submission to disease reporting systems and research infrastructures.
6. Enhance existing IT systems so that they support workflows to link and share pseudonymized data between different domains, while enabling privacy and security. Use domain specific, time stamped, encrypted person identifiers for this purpose.
7. Share Metadata where there are restrictions in accessing/using the related data.
8. Document methodologies used to collect and compile data, including data management, data cleaning, correcting errors, updating, data imputation, computer code used, definitions used, etc.
9. Provide well documented, near-real-time, tidy, easily Findable, Accessible, Interoperable, Reusable, Ethical, and Reproducible (FAIRER) data in

human- and machine-readable format (e.g, CSV or RDF), accessible on a version-controlled platform (e.g., GitHub).

10. Rapid development of standardised tools for aggregating microdata to a harmonized format(s) that can be shared and used while minimising the re-identification risk for individual records.
11. Rapid development of: (a) Resolvable Persistent Identifiers, rather than unstable Uniform Resource Locators (URLs), to provide the ability to successfully access the data over decades; (b) Machine readable citations that allow machines to access and interpret the resource; (c) Micro-citations that refer to the specific data used from large datasets; and, (d) Date and Time Access citations for dynamic data (ESIP 2019).
12. Rapid development of a new data management system infrastructure that ensures scientific data integrity via data management plans embedded in linked data life cycles that: (a) are fully machine-enabled, and not constrained by non-digital processes; (b) are available online end-to-end; (c) guarantee FAIRER data, metadata, and code/scripts; (d) guarantee data security; and, (e) provide tiered access to restricted data by appropriately credentialed users and machines.
13. Rapid development of government and institutional policies to accelerate the implementation of Open Data and Open Science tools and methods across all science and health domains.
14. Urgently update data sharing policies and Memoranda of Understanding (MOUs) across all domains, in government, healthcare systems, and research institutions to support Open Data, Open Science, scientific data modernization, and linked data life cycles that will enable rapid and credible scientific and epidemiologic discovery.
15. Emphasize that we are still in the midst of the current COVID-19 pandemic, so all data and models are therefore incomplete, provisional and subject to correction. Thus, current calculations underestimate prevalence and therefore may be underestimating and missing elements of the true risk. Nonetheless, public health experts provide the best possible scientifically based recommendations under these changing conditions and revise when necessary as understanding improves over time with the acquisition of new data.
16. Implement a policy of openness, transparency, and honesty with respect to COVID-19 related data and models, and what we know and do not know, to build and maintain public trust.

6.3 Additional Working Documents & Links

Additional materials from the Working Group can be found at: RDA-COVID19-Epidemiology

6.4 References

Battegay, M., Kuehl, R., Tschudin-Sutter, S., Hirsch, H. H., Widmer, A. F., & Neher, R. A. (2020). 2019-novel Coronavirus (2019-nCoV): Estimating the case fatality rate – a word of caution. Swiss Medical Weekly, 150(0506). https://doi.org/10.4414/smw.2020.20203

CDC. (2020a). Cases of COVID19 in the U.S. [Dataset]. Centers for Disease Control, USA. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html

CDC. (2020b). Weekly provisional death counts from death certificate data: COVID19, pneumonia, flu [Dataset]. Center for Disease Control, USA. https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm

ECDC. (2020). Geographic distribution of COVID-19 cases worldwide [Dataset]. European CDC. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

ELIXIR. (2020). ELIXIR. https://elixir-europe.org/about-us

ESIP. (2019). Data Citation Guidelines for Earth Science Data , Version 2. Earth Science Information Partners. https://doi.org/10.6084/m9.figshare.8441816.v1

European Commission. (2020). Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs. European Union. https://www.rd-alliance.org/system/files/documents/H2020_Guidelines_COVID19_EC.pdf

FAIRsharing. (2020). (a) FAIRsharing COVID-19 Collection: resources with COVID-relevant datasets: https://fairsharing.org/collection/COVID19Resources; (b) List of standards and repositories in the 'epidemiology' area: https://fairsharing.org/search/?q=epidemiology

Ghani, A. C., Donnelly, C. A., Cox, D. R., Griffin, J. T., Fraser, C., Lam, T. H., Ho, L. M., Chan, W. S., Anderson, R. M., Hedley, A. J., & Leung, G. M. (2005). Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease. American Journal of Epidemiology, 162(5), 479–486. https://doi.org/10.1093/aje/kwi230

Hausman, J., Stall, S., Gallagher, J., & Wu, M. (2019). Software and Services Citation Guidelines and Examples. ESIP. https://esip.figshare.com/articles/Software_and_Services_Citation_Guidelines_and_Examples/7640426

Johns Hopkins University. (2020). COVID19 dataset [Dataset]. Johns Hopkins University, CSSEGISandData. https://github.com/CSSEGISandData/COVID-19 (Original work published 2020)

Pathak, E. B., Salemi, J. L., Sobers, N., Menard, J., & Hambleton, I. R. (2020). COVID-19 in Children in the United States: Intensive Care Admissions, Estimated Total Infected, and Projected Numbers of Severe Pediatric Cases in 2020. Journal of Public Health Management and Practice, Publish Ahead of Print. https://doi.org/10.1097/PHH.0000000000001190

The Atlantic. (2020). The COVID Tracking Project [Dataset]. https://covidtracking.com

University of Oxford. (2020a). COVID19 dataset [Dataset]. https://github.com/owid/covid-19-data
University of Oxford. (2020b). COVID19 government response tracker [Dataset]. University of Oxford. https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker
University of Washington. (2020). COVID19 data: Beoutbreakprepared. https://github.com/beoutbreakprepared
WHO. (2020). COVID19 situation reports [Dataset]. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports
WHO, "WHO Global COVID-19 Clinical Platform Case Record Form (CRF)," Mar. 23, 2020. https://www.who.int/docs/default-source/coronaviruse/who-ncov-crf.pdf
Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L. B. da S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1–9. https://doi.org/10.1038/sdata.2016.18
World Bank. (2020). Understanding the Coronavirus (COVID-19) pandemic through data [Dataset]. http://datatopics.worldbank.org/universal-health-coverage/covid19
World Health Organization (WHO). (2020). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19).
Worldometers. (2020). COVID19 data [Dataset]. https://www.worldometers.info/coronavirus
Zhang, Y. (2020). The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19)—China, 2020. Cina CDC Weekly. http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51

7. Omics Sub-Group Guidelines

7.1 Sub-Group Focus and Description

For the purpose of this group, OMICS are defined as data from cell and molecular biology. For most of the data modalities, data can be deposited in existing deposition database resources. Many of these resources are now supporting specific COVID-19 subsets.

Within this scope, the group has prioritized recommendations on data that is already frequently associated with research on COVID-19. Additional guidance for more omics data types will be added at a later date.

7.2 Initial Sub-Group Guidelines

7.2.1 Recommendations for virus genomics data

Repositories

There are several genomics resources that can be used to make virus genomics sequences available for further research. A curated list can be found in FAIRsharing using the query "genomics" in the set of biological databases; some specific examples are listed below.

1. We suggest that **raw virus sequence data** is stored in one of these archives: INSDC: European Nucleotide Archive (ENA), DDBJ Sequence Read Archive and/or NCBI SRA. Each of these is *well known* and *openly accessible* for *immediate reuse* without undue delays.
2. For **assembled and annotated genomes** we suggest deposition in one or more of these archives: NCBI GenBank, DDBJ Annotated/Assembled sequences, European Nucleotide Archive (ENA) Assembled/Annotated sequences, and/or NCBI Virus.
3. There are other archives suitable for genome data that are more restrictive in their data access; submission to such resources is not discouraged, but such archives should not be the only place where a sequence is made available.
4. Before submission of raw sequence data from e.g. shotgun sequencing to INSDC archives, it is necessary to remove contaminating human reads.

Data and metadata formats

A list of relevant data and metadata can be found in FAIRsharing using the query 'genomics'.

1. We suggest that data is preferentially stored in the following formats, in order to maximize the interoperability with each other and with standard analysis pipelines:
   Raw sequences: fastq; optionally adding compression with gzip
   1.1. Genome contigs: fastq, fasta; optionally adding compression with gzip
      1.1.1. *De novo* aligned sequences: .afa
   1.2. Gene Structure: .gtf
   1.3. Gene Features: .gff
   1.4. Sequences mapped to a genome: .sam or the compressed formats .bam or .cram. Please ensure that the used reference sequence is also publically available and that the @SQ header is present and unambiguously describes the used reference sequence.
   1.5. Variant calling: .vcf. Please ensure that the used reference sequence is also publically available and that it is unambiguously referenced in the header of the vcf file, e.g. using the URL field of the ##contig field.

1.6. Browser: .bed

Other guidance for virus genomics

Future versions of these recommendations will add suggestions for metadata formatting, analysis pipelines, and how to reuse existing data.

7.2.2 Recommendations for host genomics data

Host genomics data is often coupled to human subjects. This comes with many ethical and legal obligations that are documented in a separate chapter and not repeated here.

Repositories

Several different types of host genomics data are collected for COVID-19 research. Some suitable repositories for these are:

1. For human transcriptomics: EGA European Genome-Phenome Archive (if the data must be be stored locally, EGA is working on a software package that can be installed locally and connects to the central metadata archive for findability); Japanese Genotype-phenotype Archive; dbGAP. A curated list of more relevant repositories can be found in FAIRsharing using the query' 'transcriptomics'.
2. Cell lines/Animals: ArrayExpress/European Nucleotide Archive (ENA), Japanese Genotype-phenotype Archive,/DDBJ Sequence Read Archive or NCBI GEO/SRA
3. GWAS: GWAS Catalog; EGA; GWAS Central
4. Gene expression: Expression Atlas, GEO, https://aoe.dbcls.jp/en, ArrayExpress
5. Adaptive Immune Receptor Repertoire sequence (AIRR-seq) annotations: AIRR Data Commons, NCBI

Data and metadata standards

A list of relevant data and metadata standards can be found in FAIRsharing using the query 'transcriptomics'

1. Preferred file formats are for transcriptomics data:
   1.1. Raw sequences: fastq (compression can be added with gzip)
   1.2. Mapped sequences: .sam (compression with .bam or .cram)
   1.3. Transcript count: TPM .csv
2. For other data types:
   2.1. GWAS: binary files (.bim .fam and .bed); text-format files (.ped and .map)
   2.2. Gene Expression: tab-delimited text, raw data file formats from commercial microarray platforms (Affymetrix, Illumina etc)
   2.3. AIRR-seq data: AIRR Repertoire metadata (json, yaml), AIRR Rearrangements (tsv)

3. Preferred minimal metadata standards are MINSEQE for Transcriptomics, MIxS for GWAS, MIAME for Gene Expression data, and MiAIRR for Immune Repertoire sequencing data.

7.2.3 Recommendations for Structural Data

A chapter on recommendations for structural data will be added to a future version of these recommendations.

7.2.4 Computing resources for COVID-19 OMICS

Several organizations providing computing resources have calls out for projects involving COVID-19 data. Some notable examples:

1. Galaxy-ELIXIR resources
2. PRACE COVID19 call
3. Exscalate 4CoV

7.3 Additional Working Documents & Links

Additional materials from the Working Group can be found at: RDA-COVID19-Omics

7.4 References

AIRR. *Community Data Standards*. 2017. Adaptive Immune Receptor Repertoire, 10 Apr. 2020. *GitHub*, https://github.com/airr-community/airr-standards.

RCSB Protein Data Bank. *RCSB Protein Data Bank SARS-CoV-2 Resources*. *www.rcsb.org*, https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true. Accessed 23 Apr. 2020.

Biocomputing Unit - Centro Nacional de Biotecnología (CNB), and Joan Segura Mora. "3DBioNotes: Automated Biochemical and Biomedical Annotations on Covid-19-Relevant 3D Structures." *3D BIONOTES-WS*, https://3dbionotes.cnb.csic.es/ws/api. Accessed 23 Apr. 2020.

COVID-19 Biohackathon April 5-11 Participants, et al. "A COVID-19-Specific Instance for EOSC-Life's WorkflowHub." *The WorkflowHub*, https://covid19.workflowhub.eu/. Accessed 23 Apr. 2020.

COVID19-hg. *COVID-19 Host Genetics Initiative*. 2020, https://www.covid19hg.org/.

DDBJ Center. "Bioinformation and DDBJ Center." *Bioinformation and DDBJ Center*. *www.ddbj.nig.ac.jp*, https://www.ddbj.nig.ac.jp/index-e.html. Accessed 24 Apr. 2020.

Di Cosmo, Roberto, et al. "Identifiers for Digital Objects: The Case of Software Source Code Preservation." *IPres2018*, ACM, 2018, p. 9.

EGA. "European Genome-Phenome Archive." *The European Genome-Phenome Archive*, https://ega-archive.org/. Accessed 17 Apr. 2020.

ELIXIR. "COVID-19: The Bio.Tools COVID-19 Coronavirus Tools List." *Bio.Tools · Bioinformatics Tools and Services Discovery Portal*. *bio.tools*, https://bio.tools/t?domain=covid-19. Accessed 23 Apr. 2020.

EMBL-EBI. *COVID 19 Data Portal*. 19 Apr. 2020, https://www.ebi.ac.uk/covid-19.

European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI). *EMBL-EBI COVID-19 Data Portal*. https://www.ebi.ac.uk/covid-19. Accessed 23 Apr. 2020.

---. “European Nucleotide Archive.” *European Nucleotide Archive < EMBL-EBI*, https://www.ebi.ac.uk/ena. Accessed 23 Apr. 2020.
FAIRsharing. “FAIRsharing.” *Oxford E-Research Centre University of Oxford*, https://fairsharing.org/. Accessed 15 Apr. 2020.
FAIRsharing Team. “IReceptor Public Archive.” *FAIRsharing*, 2020, doi:10.25504/FAIRSHARING.EKDQE5.
Galaxy Team. “COVID-19 Analysis on Usegalaxy.★.” *COVID-19 Analysis on Usegalaxy.★*, https://covid19.galaxyproject.org/. Accessed 23 Apr. 2020.
GenBank. “SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) Sequences.” *U.S. Center for Disease Control*, 2020, https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/.
Genomics Standards Consortium. “Minimum Information about Any (X) Sequence (MIxS) Specification.” *Genomics Standards Consortium / Mixs*, 1 Jan. 2019, https://github.com/GenomicsStandardsConsortium/mixs.
GitHub. “Choose an Open Source License.” *Choose a License*. *choosealicense.com*, https://choosealicense.com/. Accessed 20 Apr. 2020.
GitHub, Inc. “Making Your Code Citable.” *Making Your Code Citable; GitHub Guides*, Oct. 2016, https://guides.github.com/activities/citable-code/.
Jiménez, Rafael C., et al. “Four Simple Recommendations to Encourage Best Practices in Research Software.” *F1000Research*, vol. 6, 2017. *PubMed Central*, doi:10.12688/f1000research.11407.1.
Lamprecht, Anna-Lena, et al. “Towards FAIR Principles for Research Software.” *Data Science*, edited by Paul Groth, Nov. 2019, pp. 1–23. *DOI.org (Crossref)*, doi:10.3233/DS-190026.
LifeScience RI. *LS RI Response to COVID-19, European Life Science Research Infrastructure) | LifeScience RI*. https://lifescience-ri.eu/ls-ri-response-to-covid-19.html. Accessed 23 Apr. 2020.
Nextstrain Team, et al. *Nextstrain Genomic epidemiology of novel coronavirus*. 2020, https://nextstrain.org/ncov/global.
NIAID. *Data Sharing and Release Guidelines*. National Institute of Allergy and Infectious Disease, 17 Oct. 2013, https://www.niaid.nih.gov/research/data-sharing-and-release-guidelines.
NIAID DAIT. *ImmPort Shared Data*. 2018, https://immport.org/shared/home.
Nüst, Daniel, et al. *Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science*. OSF Preprints, Apr. 2020. *osf.io*, doi:10.31219/osf.io/fsd7t.
Opensource.org. *Licenses & Standards | Open Source Initiative*. https://opensource.org/licenses. Accessed 20 Apr. 2020.
RDA-COVID19-Omics Subgroup. “RDA-COVID19-Omics.” *RDA*, 30 Mar. 2020. *www.rd-alliance.org*, https://www.rd-alliance.org/groups/rda-covid19-omics.
Swiss Institute of Bioinformatics (SIB). “SARS-COV-2, COVID-19 Coronavirus Resource: SARS Coronavirus 2 (SARS-CoV-2) Proteome.” *SARS Coronavirus 2 ~ ViralZone Page*, https://viralzone.expasy.org/8996. Accessed 23 Apr. 2020.
The Human Protein Atlas consortium. *SARS-CoV-2 Related Proteins - The Human Protein Atlas*. https://www.proteinatlas.org/humanproteome/sars-cov-2. Accessed 23 Apr. 2020.
UniProt. “COVID-19 UniProtKB.” *UniProt*, https://covid-19.uniprot.org/uniprotkb?query=*. Accessed 23 Apr. 2020.
wwPDB consortium, et al. “Protein Data Bank: The Single Global Archive for 3D Macromolecular Structure Data.” *Nucleic Acids Research*, vol. 47, no. D1, Jan. 2019, pp. D520–28. *DOI.org (Crossref)*, doi:10.1093/nar/gky949.

8. Social Sciences Sub-Group Guidelines

8.1 Sub-Group Focus and Description

Data from the social sciences is essential for all domains (including omics, clinical, epidemiology) seeking to better plan for effective management of COVID-19 and understanding its impact. Social scientists are collecting new information and reusing existing data sources to better inform leaders and policymakers about pressing social issues regarding COVID-19, to enable evidence-based decision-making. Key data types in the social sciences include qualitative; quantitative; geospatial; audio, image, and video; and non-designed data (also referred to as digital trace data). Recommendations made in these guidelines will help ensure that data contributions from the social sciences are shared in ways that allow them to be leveraged for the broadest impact and reused across all domains.

8.2 Initial Sub-Group Guidelines

The overall principle we believe to be appropriate in times of public crises like COVID-19 is to allow the sharing of as much data as openly as possible and in a timely fashion, maintaining the public trust. We recognise, however, that this requires appropriate ethical and legal considerations. We, therefore, believe the following recommendations in relation to metadata, storage, sharing and ethical and legal issues should be referenced in making decisions which necessarily balance individual and public rights and benefits.

8.2.1 Data Management Responsibilities and Resources

1. Create a Data Management Plan (DMP) at the beginning of the research process when it can be included in the work plan and the budget and subsequently guide the handling of the data and help all disciplines understand the data. The DMP is a "living" document, which may change over the course of a project, and helps to document data for reuse and findability. Projects already underway that might contribute data to address COVID-19 should update their DMPs to ensure alignment with current recommendations.
2. When writing the DMP, contact the repository of your choice which may offer guidelines for the DMPs in advance of deposit.
3. Researchers should aim to register their DMP as an openly accessible, public deliverable.
4. Consult the European Commission Guidelines for open access to publications, data and other research outputs for Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics, and address the relevant aspects of making the data FAIR in a DMP.

8.2.2 Documentation, Standards, and Data Quality:

1. Researchers and statistical agencies should provide thorough documentation about the research context, methods used to collect data,

and quality-assurance steps taken, as well as consider the minimal number of metadata variables shared that will allow linking the different types of data produced around COVID-19.

2. Given the multidisciplinary nature of pandemic research, social scientists should provide sufficiently detailed and explicit documentation, including metadata, such that data can be understood by researchers and be machine-actionable to allow for broad and interdisciplinary use. To do so, researchers should utilise community-endorsed metadata standards, controlled vocabularies and ontologies, and recommended file formats.
3. To encourage interdisciplinary research, social scientists should be mindful of commonly accepted professional codes or norms for documentation needs when producing documentation according to their own particular disciplinary norms. This allows for all domains to be able to ensure the research integrity of social science data it accesses or reuses.

8.2.3 Storage and Backup:

1. Research institutions should provide researchers with robust data storage facilities that follow recommendations regarding areas such as regular backup in multiple locations and data protection. Where possible, researchers should use the official storage provisions available from their institution, including when working remotely.
2. Data may have particular requirements as to how it can be stored and accessed, based on laws and regulations, research ethics protocols, or secondary data licenses. Sensitive data and human subject data containing personally identifiable information (PII) or protected health information (PHI) should be adequately protected and encrypted when at rest or in transit. Where possible, store sensitive data without direct identifiers.

8.2.4 Legal and Ethical Requirements:

1. Find a balance that takes into account individual, community and societal interests and benefits whilst addressing public health concerns and objectives to enable access to data and their reuse, and maximise the research potential.
2. It is recommended to establish rigorous approval mechanisms for sharing data (via consent, regulation, institutional agreements and other systematic data governance mechanisms). Researchers have a responsibility for ensuring research participants understand that there may be a risk of re-identification when data are shared.
3. Ethics review during a crisis like the COVID-19 pandemic is critical to protect highly vulnerable populations from potential harm. Therefore this report endorses guidance such as the Statement of the African Academy of Sciences' Biospecimens and Data Governance Committee On COVID-19: Ethics, Governance and Community engagement in times of crises.
4. Where possible, provide immediate open access to all relevant research data. Open data should be licensed under Creative Commons Attribution 4.0 International License (CC BY 4.0) or a Creative Commons Public Domain Dedication (CC0 1.0) or equivalent. If immediate open access is not possible, researchers should make data available as soon as possible. Researchers whose data have legal, privacy, or other restrictions should seek out appropriate alternative avenues for data sharing, including restricted access conditions.
5. Ensure licenses and agreements in data acquisition enable downstream data sharing and preservation. If working with commercial partners, seek

opportunities to negotiate data sharing mechanisms agreeable and equitable to all parties.

8.2.5 Data Sharing and Long-term Preservation:

1. Ensure data shared is FAIR: Findable, Accessible, Interoperable and Reusable. In the current emergency context, it is a moral imperative to share the data and preserve it in the most open way possible for each case.
2. Select data for long term preservation; researchers should retain data that underpin published findings, data that allow for validation and replication of results, and the broader set of data with long-term value.
3. Deposit quality-controlled research data in a data repository, whenever possible in a trustworthy digital repository committed to preservation, such as one having undergone formal certification. As the first choice, disciplinary repositories are recommended for maximum visibility, followed by general or institutional repositories.
4. In order to expedite re-use, data that could be used to advance research on pandemics should be given top priority in the data publication process, fast-tracked by repositories, institutions, and other data publishers.
5. To ensure social sciences data can be linked with data being produced by other entities, consider preserving information that enables data linkages to be made, under appropriate security frameworks by creating a separate file, to be kept apart from the rest of the data, which provides the linking relationship between any personal identifiers and the randomly assigned unique identifiers.
6. Repositories should provide key metadata associated with its datasets, optimally utilising a metadata standard that allows for interoperability. They also should employ tools such as persistent identifiers for discovering and citing the data, as well as mechanisms for linking data and other research objects.
7. Researchers should make available and deposit with data in a repository all documentation--such as codebooks, lab journals, informed consent form templates--which are important for understanding the data and combining them with other data sources. Researchers should also make available information regarding the computing context relevant for using the data (e.g., software, hardware configurations, syntax queries) and deposit with the data where possible.

8.3 Additional Working Documents & Links

For the complete guidelines please see RDA COVID-19 WG Guidelines - Social Sciences

Additional materials from the Working Group can be found at: RDA-COVID19-Social-Sciences

9. Overarching Research Software Guidelines

This contribution to the RDA COVID-19 Guidelines was provided by the Research Software Alliance <ReSA>.

9.1 Focus and Description

It is important to put forward some key practices for the development and (re)use of research software, as these facilitate code sharing and accelerated results in responses to the COVID-19 pandemic. This section will be relevant to audiences ranging from researchers and research software engineers with comparatively high levels of knowledge about software development to experimentalists, such as wet-lab researchers, with almost no background in software development.

Seven clear, practical recommendations around basic software principles and practices are provided here, in order to facilitate the open and clear collaborations that can contribute to resolving current challenges. These recommendations aim to enable relatively small points of improvement across all aspects of software that will allow its swift (re)use, facilitating the accelerated and reproducible research needed during this crisis. These recommendations highlight key points derived from a wide range of work on how to improve your research software right now, to achieve better research (Wilson et al. 2017, Jiménez et al. 2017, Lamprecht et al. 2019; Akhmerov et al. 2019; Clément-Fontaine et al. 2019).

9.2 Initial Guidelines

1. Make your software available: Making software that has been developed available is essential for understanding your work, allowing others to check if there are errors in the software, be able to reproduce your work, and ultimately, build upon your work. The key point here is to ensure that the code itself is shared and freely available (see about licenses below), through a platform that supports access to it and allows you to effectively track development with versioning (e.g. code repositories such as GitHub, Bitbucket, GitLab, etc.)
    1.1. Resources:
        1.1.1. Four Simple Recommendations to Encourage Best Practices in Research Software https://doi.org/10/gbp2wh
        1.1.2. FAIR Software guidelines on code repositories https://fair-software.nl/recommendations/repository
2. Reference your software with Persistent Identifiers (PIDs): Equally important to making the source code available is providing a means of referring to it (Cosmo et al. 2018). For this reason, software should be deposited within a repository that supports persistent identifiers (PIDs - a specific example being DOIs) such as Zenodo, Figshare or Software Heritage which provides more persistent storage than the above code repositories in R1.
    2.1. Resources:
        2.1.1. FAIR software guidelines on citing software https://fair-software.nl/recommendations/citation
        2.1.2. List of software registries https://github.com/NLeSC/awesome-research-software-registries
        2.1.3. Making your code citable through GitHub and Zenodo

https://guides.github.com/activities/citable-code/

3. Provide metadata/documentation that describe at least the libraries and parameters used: (Re)using code/software requires knowledge of two main aspects at minimum: environment and expected input/output. The goal is to provide sufficient information that computational results can be reproduced and may require a minimum working example.
    3.1. Resources:
        3.1.1. Ten simple rules for documenting scientific software https://doi.org/10.1371/journal.pcbi.1006561
4. Ensure portability and reproducibility of results: It is critical, especially in a crisis, for software that is used in data analysis to produce results that can, if necessary, be reproduced. This requires automatic logging of all parameter values (including setting random seeds to predetermined values), as well as establishing the requirements in the environment (dependencies, etc). Container systems such as Docker or Singularity can replicate the exact environment for others to run software/code in.
    4.1. Resources:
        4.1.1. Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science https://doi.org/10.31219/osf.io/fsd7t.
5. Release your software under a licence: Software code is typically protected by copyright in most countries, with copyright often held by the institution that does the work rather than the developer themself. By providing a licence for your software, you grant others certain freedoms, i.e. you define what they are allowed to do with your code. Free and Open Software licenses typically allow the user to use, study, improve and share your code.
    5.1. Resources:
        5.1.1. Choose an Open Source License https://choosealicense.com/.
6. Respect the licences of the software you have used: The freedom provided by your license must not violate the restriction described in the license of the third-party dependency that you are using. If you are using commercial software then it is likely you will not be able to share the package freely, but this depends of course on the licence terms.
    6.1. Resources:
        6.1.1. 4 Tips for Keeping on Top of Project Dependencies https://osswatch.jiscinvolve.org/wp/2013/05/08/4-tips-for-keeping-on-top-of-project-dependencies/
7. Cite the software you use: It is good practice to acknowledge and cite the software you use in the same fashion as you cite papers to both identify the software and to give credit to its developers. For software developed in an academic session, this is the most effective way of supporting its continued development and maintenance because it matches the current incentives of that system.
    7.1. Resources:
        7.1.1. Software Citation Principles https://doi.org/10.7717/peerj-cs.86.

9.3 References

Akhmerov, Anton, Maria Cruz, Niels Drost, Cees Hof, Tomas Knapen, Mateusz Kuzak, Carlos Martinez-Ortiz, Yasemin Turkyilmaz-van der Velden, and Ben van Werkhoven. 2019. "Raising the Profile of Research Software." Zenodo. https://doi.org/10.5281/zenodo.3378572.

Awesome list of Research Software Registries, https://github.com/NLeSC/awesome-research-software-registries.

Choose an Open Source License - https://choosealicense.com/.

Clément-Fontaine, Mélanie, Roberto Di Cosmo, Bastien Guerry, Patrick MOREAU, and François Pellegrini. 2019. "Encouraging a Wider Usage of Software Derived from Research." Research Report. Committee for Open Science's Free Software and Open Source Project Group. https://hal.archives-ouvertes.fr/hal-02545142.

FAIR Software - Five Recommendations for FAIR Software, https://fair-software.nl/.

GitHub Guides - Making your code citable, https://guides.github.com/activities/citable-code/.

Jiménez, Rafael C., Mateusz Kuzak, Monther Alhamdoosh, Michelle Barker, Bérénice Batut, Mikael Borg, Salvador Capella-Gutierrez, et al. 2017. "Four Simple Recommendations to Encourage Best Practices in Research Software [Version 1; Referees: 3 Approved]." *F1000Research* 6 (June): 876. https://doi.org/10/gbp2wh.

Lamprecht, Anna-Lena, Leyla Garcia, Mateusz Kuzak, Carlos Martinez, Ricardo Arcila, Eva Martin Del Pico, Victoria Dominguez Del Angel, et al. 2019. "Towards FAIR Principles for Research Software." Edited by Paul Groth. *Data Science*, November, 1–23. https://doi.org/10.3233/DS-190026.

Lee BD Ten simple rules for documenting scientific software. PLoS Comput Biol 14(12): e1006561. 2018. https://doi.org/10.1371/journal.pcbi.1006561

Nüst, Daniel, Vanessa Sochat, Ben Marwick, Stephen Eglen, Tim Head, and Tony Hirst. 2020. "Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science," April. https://doi.org/10.31219/osf.io/fsd7t.

Smith, Arfon M., Daniel S. Katz, Kyle E. Niemeyer, and FORCE11 Software Citation Working Group. 2016. "Software Citation Principles." *PeerJ Computer Science* 2 (September): e86. https://doi.org/10.7717/peerj-cs.86.

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. 2017. "Good Enough Practices in Scientific Computing." *PLOS Computational Biology* 13 (6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510.

Wilson, Scott. 4 Tips for Keeping on Top of Project Dependencies, https://osswatch.jiscinvolve.org/wp/2013/05/08/4-tips-for-keeping-on-top-of-project-dependencies/

10. Overarching Legal and Ethical Guidelines

The second cross-cutting topic to be explored in depth, is Legal and Ethical. This sub-group is currently in formation, and will have a draft output in the coming weeks.