RDA COVID-19 Working Group

Recommendations and Guidelines

2nd release

1 May 2020

RDA Recommendation (2nd Release - 1 May 2020)
Produced by: RDA COVID-19 Working Groups, 2020

Document Metadata

Identifier DOI: 10.15497/RDA00046

Title RDA COVID-19; recommendations and guidelines, 2nd release 1 May 2020

Description This is the second 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 themes of Research Software andLegal and Ethical .

Date Issued 2020-05-01

Version Draft guidelines and recommendations; second release, 1 May 2020, version for public review

Contributors RDA COVID-19 Working Groups

Group Co-chairs: *Juan Bicarregui, Anne Cambon-Thomsen, Ingrid Dillo, Natalie Harrower, Sarah Jones, Mark Leggott, Priyanka Pillai*

Subgroup Moderators:
Clinical: Sergio Bonini, Dawei Lin, Andrea Jackson-Dipina
Community Participation: Timea Biro, Kheeran Dharmawardena, Eva Méndez, Daniel Mietchen, Susanna Sansone, Joanne Stocks
Epidemiology: Claire Austin, Gabriel Turinici
Legal and Ethical: Alexander Bernier, John Brian Pickering
Omics: Natalie Meyers, Rob Hooft
Social Sciences: Iryna Kuchma, Amy Pienta
Software: Michelle Barker, Hugh Shanahan, Fotis Psomopoulos

Editorial team: *Christoph Bahim, Alexandre Beaufays, Ingrid Dillo, Natalie Harrower, Mark Leggott, Nicolas Loozen, Priyanka Pillai, Mary Uhlmansiek, Meghan Underwood, Bridget Walker*

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 'RDA-COVID19,' 'RDA-COVID19-Clinical,' 'RDA-COVID19-Community-participation,' 'RDA-COVID19-Epidemiology,' 'RDA-COVID19-Legal-Ethical,' 'RDA-COVID19-Omics,' 'RDA-COVID19-Social-Sciences,' 'RDA-COVID19-Software,' and we acknowledge the support provided by the RDA community and structure.

Table of Contents

0. Log Changes

Document	**Changes**
RDA COVID-19; recommendations and guidelines, 1st release 24 April 2020	First draft of document released for comments and feedback
RDA COVID-19; recommendations and guidelines, 2nd release 1 May 2020	Section 3 -Foundational Principles/Recommendations - modifications Section 4 - Clinical - updated Section 6 - Epidemiology - revised Section 8 - Omics - revised Section 9 - Overarching Research Software Guidelines - new sub-section 9.3 initial guidelines for policy makers included Section 10 - Overarching Legal and Ethical Guidelines added Incorporation of feedback received via Requests for Comments (RfC) process and directly to Co-chairs, moderators and editorial team

1. Function of the RDA COVID-19 WG

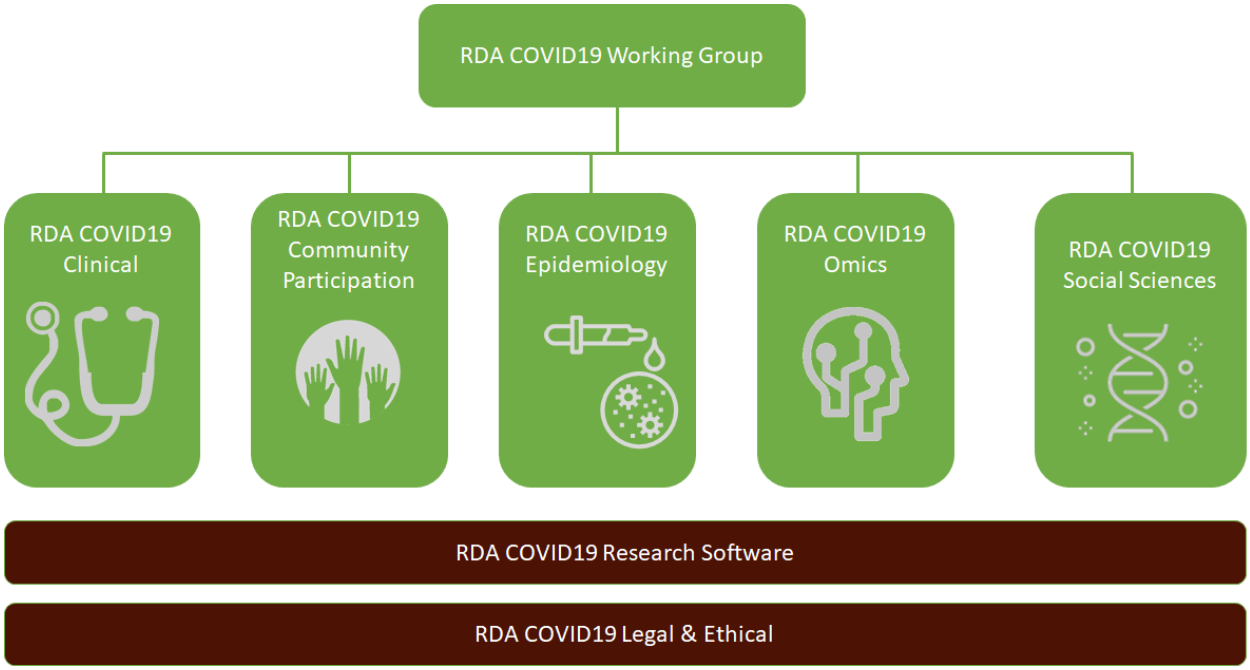
During a pandemic, data combined with the right context and meaning can be transformed into knowledge for informing public health response. Timely and accurate collection, reporting and sharing of data with the research community, public health practitioners, clinicians and policy makers will inform 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, evidenced by the 117 cross-sectoral signatories to the Wellcome Trust statement on 31st January 2020, and the further agreement by 30 leading publishers on immediate open access to COVID-19 publications and underlying data.

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, quality 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 others while providing a tool which could serve as the standard of *good enough to decide*. 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 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 May 1, there were well over 440 members of the CWG, relatively evenly spread across the 5 groups, as well as with the addition of new cross-cutting groups. A preliminary draft was released April 24th indicating early progress in the space of 3 weeks, and was opened for comment. 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. The current timeline will produce a new release each Friday, with a penultimate release anticipated for May 25, followed by a round of community feedback, and a final release of Version 1 of the Guidelines for June 15. These 5 thematic guidelines will be enhanced with a series of "cross-cutting" guidelines, that articulate principles and recommendations common across all 5 areas. Two initial cross cutting guidelines are included here: Research Software and Legal and Ethical. Additional cross-cutting guidelines suggestions are welcome. This effort also reflects the work of a host of other RDA Working Groups, as well as external stakeholder organizations, that has developed over a number of years - we want to recognize and highlight those efforts.

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: Speed vs. Accuracy

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. Raw data, algorithms, workflows, models, software and so on are required inputs to research studies, and are essential to the scientific discovery process itself. 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.

The inaccuracy and/ low quality of data shared within such short timelines, could have considerable implications, for example:

1) shortcuts with the interpretation of data can create issues, such as the debate on whether COVID-19 is 'just another flu' or not;
2) obligation to share data could orient at least some institutions to reduce testing (suspected cases do not, count, only confirmed ones do, and hence lowering testing allows lowering confirmed case numbers and creates the illusion that the epidemic is under control);
3) in some cases, lack of transparency and publication of false numbers is perhaps worse than no publication at all.

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.

The challenge here is the trade off between timeliness and precision. The speed of data collection and sharing needs to be balanced with accuracy, which takes time. The pressure to interpret results, turn around studies quickly and update statistics in almost real-time must not compromise quality and reliability. There is no overarching formula for finding that balance, but documented transparency in the research process and decisions taken can help to mitigate the dangers associated with working at hyperspeed.

3.1.3 Lack of Coordinated Standards and Context

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.

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. For a registry of metadata standards, see RDA's Metadata Standards Directory. 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. In general, you can consult re3data.org for a searchable database of research data repositories. Repositories certified by CoreTrustSeal, a result OF the RDA Repository Audit and Certification DSA–WDS Partnership WG and the European Commission's expert group on FAIR data are listed here.

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. While the privacy protection and anonymisation challenges are substantial (ie. as evidenced with current discussions about contact tracing) solutions which allow algorithms to 'visit' data, asking specific research questions which can be answered while not allowing direct access to data, should be considered.

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 actions were considered in the scope of the sub-group, the work of the Clinical Subgroup centers first on dealing with consent on data sharing, how clinical trials are conducted, how clinical information (personal and health data) and results are shared and consumed in a trustworthy and efficient manner.

4.2 Initial Sub-Group Guidelines

4.2.1 Consent on COVID-19

1. Procedures on data sharing specific for COVID-19 in the general consent for clinical trials should be in accordance with ISO/TS 17975:2015 (Health informatics — Principles and data requirements for consent in the Collection, Use or Disclosure of personal health information)

4.2.2 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 and legal principles. Specific recommendations on registering, performing, and sharing ongoing clinical research are the following:

1. Lawful fast track approval procedures of clinical trials in cases of public health emergencies exist that speed up processes while protecting adequately 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.
2. 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 efforts.
3. Multi-centres multi-countries studies including a sample size calculation according to the primary objective should be recommended to generate sound evidence on COVID-19 treatments. Collaborative trials and multi-arms studies comparing different drugs are advisable.
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) and adhere to FAIR data principles.

6. 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.

4.2.3 Sharing Clinical Data

1. In order to expedite the process of data sharing, standardized agreements for sharing of data from COVID-19 clinical trials should be developed and implemented (e.g. data transfer agreements, data use agreements).
2. Adequate tools should be implemented for collection and analysis of reliable real-world data of drugs approved for the treatment of COVID-19.
3. 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.
4. To prepare data for sharing clinical trial data should always be associated with adequate and standardised metadata to improve discoverability ("F" in FAIR).
5. Data and trial documents should be made available for sharing.
    - 5.1. 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.
        - 5.1.1. Resource: Sharing and reuse of individual participant data from clinical trials: principles and recommendations
    - 5.2. Data and trial documents should be transferred to a suitable and secure data repository (see *Trustworthy Sources of Clinical Data*) 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).
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)
    - 6.1. Preprint publishing and other forms of knowledge sharing and exchange are also encouraged.
    - 6.2. Where possible, open source journals and adherence to OpenAire initiatives and the likes are also encouraged.

More information is at clinical trials and clinical aspects documentation.

4.2.4 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: RDA-COVID19-Clinical.

4.3 Additional Working Documents & Links

1. https://clinicaltrials.gov/
2. https://www.nlm.nih.gov/NIHbmic/index.html
3. https://datascience.nih.gov/covid-19-open-access-resources
4. Sharing and reuse of individual participant data from clinical trials: principles and recommendations
5. CDISC interim User Guide for COVID-19
6. International COVID-19 Clinical Trials Map (based on the WHO Clinical Trials Search Portal): https://covid-19.heigit.org/clinical_trials.html
7. ISO/TS 17975:2015: https://www.iso.org/standard/61186.html
8. Official CDC guidelines for the new COVID-19 ICD-10-CM code: https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf

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.

We are still in the midst of the current COVID-19 pandemic. All data and models are therefore incomplete, provisional and subject to correction under changing conditions. Understanding will improve over time with the acquisition of new data. However, 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 or lack of metadata, lack of version control, fragmented datasets, delays in releasing data, non-standard definitions and reporting parameters, unavailable or undocumented computer code, copyright and usage conditions, translation requirements, consents, approvals, and legal restrictions. In addition, clinical, eHealth, surveillance, and research systems within and across jurisdictions do not integrate well, or at all.

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 insights and understanding of the newly emergent SARS-CoV-2 virus and the 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

Policy recommendations

1. 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, and to fast-track decision-making. For example, between the countries and the WHO, between the European Commission and the USA, and between sub-national jurisdictions/institutions and their national government.

2. Implement a "data first" policy by treating publication of data articles in "open" peer-reviewed data journals, including deposit of the data and associated code in a trusted digital repository, as first-class research outputs equal in value to traditional peer-reviewed articles.
3. 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.
4. Call upon the international Open Government Partnership (OGP) to add "Open Science" as one of its Policy Areas to be included in National Action Plans. Member countries would then be held accountable for developing and implementing Open Science commitments via the Independent Reporting Mechanism (IRM) that tracks the progress of OGP members.
5. Build and maintain public trust: 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. Publish situational data, analytical models, scientific findings, and reports used in decision-making and justification of decisions (OGP 2020).

IT and data management infrastructure recommendations

1. Investment in information technology (IT) and data management system infrastructure (devices or hardware, and algorithms or software used to store, retrieve and process data).
    1.1. Rapid development of a modern data management system infrastructure will ensure 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) enable synchronous and asynchronous workflows; (d) guarantee tidy, Findable, Accessible, Interoperable, Reusable, Ethical, and Reproducible (FAIRER) data, metadata, and code/scripts; (e) guarantee data security; (f) provide tiered access to restricted data by appropriately credentialed users and machines; and, (g) analytical tools. See, for example, ELIXIR Galaxy.
    1.2. When evaluating apps consider the many underlying issues: legal, confidentiality, data completeness, representativeness, data quality, reliability, verifiability, data ownership, data access, data openness, data control, transparency, peer-review, etc.

Analysis and modeling recommendations

1. Develop and implement internationally harmonized COVID-19 intervention policies based on peer-reviewed empirical modeling and epidemiological evidence.
2. Account for public health decision making in modelling COVID-19 inputs and outputs.
3. Harmonize approaches to comparably quantify side-effects of pandemic mitigation measures on society, for example, shifts in morbidity, mortality, health care utilization, quality of life, social isolation.
4. Provide uncertainty quantification of all reported parameters and conclusions for all model predictions, data etc.

5. Implement a data driven approach to identify hotspots.

Surveillance data recommendations

1. Rapid development of a consensus standard on COVID-19 surveillance data:
    1.1. Definition of and criteria for COVID-19 testing, and reporting on testing.
    1.2. 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, follow up to identify serious long-term effects in recovered cases, sequelae and immunity, location, demographic, socioeconomic information, and outcome of resolved cases.
    1.3. Daily reporting cut-off time.
2. Document methodologies used to collect and compile data, including data management, data cleaning, data quality checks, updating, data imputation, computer code used, definitions used, etc.
3. Rapid development of standardized tools for aggregating microdata to a harmonized format(s) that can be shared and used while minimizing the re-identification risk for individual records.
4. Rapid development of: (a) Resolvable Persistent Identifiers, rather than 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).

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 do not 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.

Conceptualization of an epidemiological surveillance data model (Figure 1) identifies the primary data domains that need to be integrated to understand COVID-19, and to improve surveillance and follow-up: (a) clinical event history and disease milestones; (b) epidemiological indicators and reporting data; (c) contact tracing; (d) person risk factors.

However, standardization challenges within each of these domains remain to be solved before data can be effectively integrated across domains. For example, on the clinical side, the U.S. Clinical Data Interchange Standards Consortium (CDISC) new specification (Interim User Guide for COVID-19), and the WHO Core and Rapid COVID-19 Case Reporting Forms used in low- and middle-income Countries (LMIC) require additional harmonization. Event history is a boundary with clinical data. In addition to standard treatments, such providing oxygen passively or aggressively to lungs, dialysis for kidney damage, managing coagulopathy/stroke/heart attack/pericarditis, etc., there is a capacity

concern including drug availability. In addition, compassionate or experimental treatments and trials include, for example, extracorporeal oxygenation, and ad hoc drug treatments with limited evidence of outcomes and with little potential to learn from experience. Contact history and location is an unsettled domain given community surveillance in LMICs and elsewhere, and competing visions in the rapid emergence of various apps from the academic, government, and private sectors which may or may not provide an individual's geospatial location. There is inconsistent collection of person risk factor information. New York State is developing a COVID-19 risk matrix for establishments that they plan to use in "reopening" the state. Some of the person risk factors may be interrelated, and it will take some future data science to reduce the dimensionality of the risk factor space.

Interoperability and data exchange recommendations

1. 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.
2. Develop systems that 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.
3. Share Metadata where there are restrictions in accessing/using the related data.

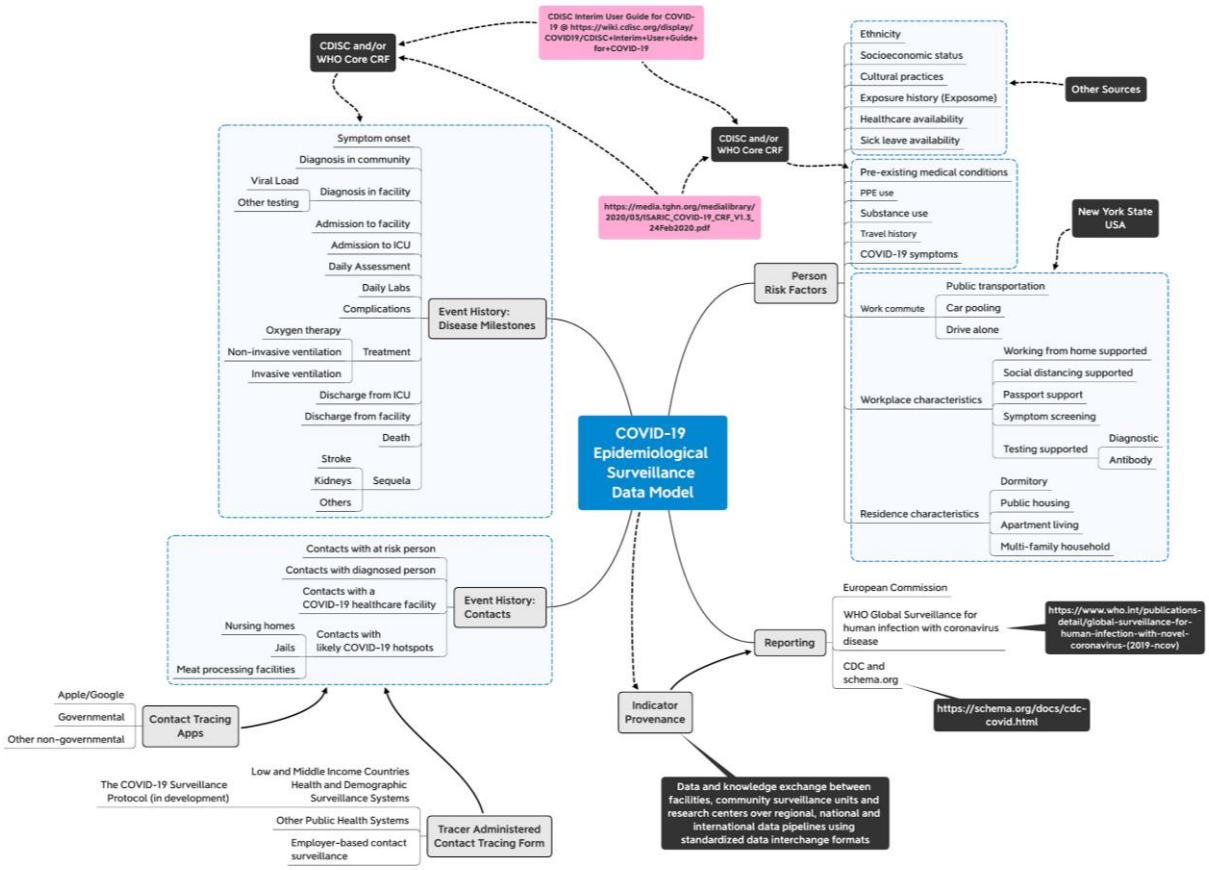


Figure 1. Epidemiology surveillance data model (draft)

Semantic interoperability is key to sharing and linking data within and across organizations. Patient related data related to the COVID-19 pandemic is handled across clinical, community surveillance (demographic and epidemiological), research, disease management, and social domains. Data in the clinical and community domains support care to patients. Regional and national administrations use some of the clinical data for disease management, e.g. in local outbreaks. Researchers generate new knowledge. Contact tracing, telemonitoring, social media are used in the social domain. In order to optimize the outcome, the data flows within and between these domains needs to be further developed to enable secure, safe, timely and reliable automated data processing.

Translational research is already leveraging existing platforms in an impromptu fashion (Westfall et al. 2007). These platforms exchange data and knowledge between facilities, community surveillance units and research centers over regional, national and international data pipelines/networks using standardized data interchange formats. However, these arrangements are developing in an ad hoc fashion and need to be evaluated to determine if they are fit-for-purpose.

In order to accomplish sustainable results, existing programs and initiatives must begin with a well-defined set of high priority short term goals, e.g. optimizing data use for disease management. In the disease management domain these include the WHO, ECDC, Tessy, Austrian EMS, CDC, NIH and the FDA. Clinical data exchange systems e.g. in the EU and USA need to be considered. In the research domain e.g. CDISC must be considered, as well as other technical standards from the more clinical space, e.g. IHE and HL7. In parallel to short term activities, long-term cooperation needs to be established, under clear coordination and with sufficient resources.

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

Brickley D (2020), "Schema.org - COVID Hospital data schema (US CDC)," Feb. 04, 2020. https://schema.org/docs/cdc-covid.html

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

CDC (2020c, March 24). Public Health and Promoting Interoperability Programs (formerly, known as Electronic Health Records Meaningful Use). https://www.cdc.gov/ehrmeaningfuluse/introduction.html

CDISC (n.d.). CDISC Standards in the Clinical Research Process [Text]. CDISC. https://www.cdisc.org/standards

Davis, L (2020). Corona Data Scraper [HTML]. COVID Atlas. https://github.com/covidatlas/coronadatascraper

DDI Alliance (2020). Data Documentation Initiative. https://ddialliance.org/
DeBord, D. G., Carreon, T., & Lentz, T (2016). Use of the "Exposome" in the Practice of Epidemiology: A Primer on -Omic Technologies. Am J Epidemiology, 184(4), 312–314. https://doi.org/10.1093/aje/kwv325
DSI, D. S., CEF eHealth (2020, March 24). EHDSI INTEROPERABILITY SPECIFICATIONS, Requirements and Frameworks (normative artefacts)—EHealth DSI Operations—CEF Digital. https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=35210463
ECDC (2019). The European Surveillance System (TESSy). European Centre for Disease Prevention and Control. https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy
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
EU eHealth Network (2019). EHealth Network Guidelines to the EU Member States and the European Commission on an interoperable eco-system for digital health and investment programmes for a new/updated generation of digital infrastructure in Europe, https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20190611_co922_en.pdf
European Commission (2016, November 25). European Reference Networks [Text]. Public Health - European Commission. https://ec.europa.eu/health/ern/networks_en
European Commission (2020a). Pseudonymisation tool. EUPID - European Platform on Rare Disease Registration. https://eu-rd-platform.jrc.ec.europa.eu
European Commission (2020b) Commission recommendation (EU) 2020/518 on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymized mobility data. https://ec.europa.eu/info/sites/info/files/recommendation_on_apps_for_contact_tracing_4.pdf
European Commission. (2020c). 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
European Commission (2020d). eHDSI Interoperability Specifications, Requirements and Frameworks (normative artefacts) - eHealth DSI Operations - CEF Digital, Mar. 24, 2020. https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=35210463
European Parliament and Council, Regulation (EC) No 851/2004 of the European Parliament and of the Council of 21 april 2004 establishing a European Centre for disease prevention and control, EUR-Lex - 32004R0851 - EN - EUR-Lex. 2004. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32004R0851
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
FDA Sentinel Initiative (2019). Sentinel Common Data Model | Sentinel Initiative. https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model
Finnie, T., South, A., & Bento, A (2016). EpiJSON: A unified data-format for epidemiology. Epidemics, 15(June, 2016), 20–26. https://doi.org/10.1016/j.epidem.2015.12.002

FitzHenry, F., Resnic, F. S., Robbins, S. L., Denton, J., Nookala, L., Meeker, D., Ohno-Machado, L., & Matheny, M. E (2015). Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. Applied Clinical Informatics, 6(3), 536–547. https://doi.org/10.4338/ACI-2014-12-CR-0121

FMSA, "Implementierungsleitfaden Meldung an das Epidemiologische Meldesystem (EMS)–Labor-und ArztmeldungZur Anwendung im österreichischen Gesundheitswesen [1.2.40.0.34.7.9.2.0]," Federal Ministry of Social Affairs, Health, Care and Consumer Protection, 2020. https://hl7.at/wp-content/uploads/2020/04/Implementierungsleitfaden-Meldepflichtige-Krankheiten_v02.20.pdf

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

GIDA (2019). Global Indigenous Data Alliance, https://www.gida-global.org

Google. (2020). Google Dataset Search. https://datasetsearch.research.google.com/

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

Healthcare Systems Research Network (2019). VDW Data Model. http://www.hcsrn.org/en/Tools%20&%20Materials/VDW/

Healy, Kieran (2020). Rpackage—COVID19 Case and Mortality Time Series (Version version 0.1.0) [R package]. https://kjhealy.github.io/covdata

HL7 (2020). HL7 Standards Product Brief—CDA® Release 2 | HL7 International. Retrieved April 21, 2020, from http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7

HL7 (). "HL7 FHIR® Implementation Guide: Electronic Case Reporting (eCR) - US Realm." http://hl7.org/fhir/us/ecr/index.html

HL7 (). "HL7 Standards Product Brief - CDA® Release 2 | HL7 International." http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7

HL7 (2019). "Fast Healthcare Interoperability Resources Standard," HL7®, Nov. 01, 2019. https://www.hl7.org/fhir/

IHE International (2019). "IHE Patient Care Device (PCD) Technical Framework, Volume 1, IHE_PCD_TF_Vol1.pdf," Dec. 12, 2019. https://www.ihe.net/uploadedFiles/Documents/PCD/IHE_PCD_TF_Vol1.pdf

IHE International (2019). "IT Infrastructure (ITI) Technical Framework," Integrating Healthc. Enterp., vol. 1, p. 334, 2019. https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol1.pdf

IHME (2020). COVID-19 Projections. Institute for Health Metrics and Evaluation. University of Washington, Institute for Health Metrics and Evaluation https://covid19.healthdata.org/projections

IHME (2020). Global Health Data Exchange | GHDx. University of Washington, Institute for Health Metrics and Evaluation (IHME), http://ghdx.healthdata.org/

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

Knight, G., Dharan, N., & Fox, G (2016). Bridging the gap between evidence and policy for infectious diseases: How models can aid public health decision-making. Int J Infect Dis., 42, 17–23. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4996966/

National Institutes of Health, U.S. Department of Health and Human Services. (n.d.). Open-Access Data and Computational Resources to Address COVID-19. Open-Access Data and Computational Resources to Address COVID-19 | Data Science at NIH. Retrieved April 17, 2020, from https://datascience.nih.gov/covid-19-open-access-resources

New York Times (2020). Coronavirus (Covid-19) Data in the United States. https://github.com/nytimes/covid-19-data

Nishiura, H (2017). Realizing Policymaking Process of Infectious Disease Control using Mathematical Modeling Techniques—R&D Projects: R&D Projects: R&D Projects Selected in FY2012- R&D Program: Science of Science, Technology and Innovation Policy. https://www.jst.go.jp/ristex/stipolicy/en/project/project20.html

Observational Health Data Sciences and Informatics (2019). OMOP Common Data Model – OHDSI. https://www.ohdsi.org/data-standardization/the-common-data-model/

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

Patient-Centered Outcomes Research Institute (2020). PCORnet. The National Patient-Centered Clinical Research Network. https://pcornet.org/

PEPP-PT (2020). "Pan-European Privacy-Preserving Proximity Tracing." https://www.pepp-pt.org

"European Rare Disease Registry Infrastructure (ERDRI) | EU RD Platform," Apr. 13, 2020. https://eu-rd-platform.jrc.ec.europa.eu/erdri-description_en

Semantic Scholar (2020). CORD-19. https://pages.semanticscholar.org/coronavirus-research

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

Translational Cancer Research Network, "Translational Research – Defining the 'T's'", http://www.tcrn.unsw.edu.au/translational-research-definitions

U.N (2018). The Humanitarian Exchange Language (HXL). United Nations, Office for the Coordination of Humanitarian Affairs (OCHA), Centre for Humanitarian Data. https://hxlstandard.org/standard/1-1final/

U.N (2020). The Humanitarian Data Exchange (HDX). United Nations, Office for the Coordination of Humanitarian Affairs (OCHA), Centre for Humanitarian Data. https://data.humdata.org/

University of Maryland (2020, April 24). University of Maryland COVID-19 Impact Analysis Platform. COVID-19 Impact Analysis Platform. https://data.covid.umd.edu/

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

Westfall, J. M., Mold, J., & Fagnan, L. (2007). Practice-Based Research—"Blue Highways" on the NIH Roadmap. JAMA, 297(4), 403–406. https://doi.org/10.1001/jama.297.4.403

WHO (2020a). Novel Coronavirus (2019-nCoV) situation reports. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

WHO (2020b, February). COVID-19 CRF • ISARIC. https://isaric.tghn.org/COVID-19-CRF/

WHO (2020c). Global surveillance for COVID-19 caused by human infection with COVID-19 virus: Interim guidance (WHO/2019-nCoV/SurveillanceGuidance/2020.6; p. 4). World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/331506/WHO-2019-nCoV-SurveillanceGuidance-2020.6-eng.pdf

WHO (2020d, March 23). WHO Global COVID-19 Clinical Platform Case Record Form (CRF). https://www.who.int/publications-detail/global-covid-19-clinical-platform-novel-coronavius-(-covid-19)-rapid-version

WHO (2020e, March 29). Modes of transmission of virus causing COVID-19: Implications for IPC precaution recommendations. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations
WHO (2020f). Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)
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/
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.

7.2 Initial Sub-Group Guidelines

7.2.1 Recommendations for virus genomics data

Archives

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) [submission docs], DDBJ Sequence Read Archive and/or NCBI SRA [submission docs]. Each of these is *well known* and *openly accessible* for *immediate reuse* without undue delays. Data submitted to either of these resources will be available through each of them.
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) [submission docs] Assembled/Annotated sequences, and/or NCBI Virus [submission docs].

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'; some specific examples are below.

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:
    - 1.1. Raw sequences: fastq; optionally add compression with gzip
    - 1.2. Genome contigs: fastq if uncertainties of the assembler can be captured, fasta otherwise; optionally add compression with gzip
        - 1.2.1. *De novo* aligned sequences: .afa
    - 1.3. Gene Structure: .gtf
    - 1.4. Gene Features: .gff
    - 1.5. 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.
    - 1.6. 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.
    - 1.7. Browser: .bed

Guidelines for data analysis

1. Standard genome variation software can be used. A good example is the Galaxy platform (and associated training materials).

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.

Generic recommendations for researchers producing data

Data sharing of not only summary statistics (or significant data) but also raw data (individual-level data)will foster a build-up of larger datasets. This will eventually allow identifying the determinants of phenotype more accurately.
Especially for raw sequencing data make sure to include QC results and details of the sequencing platform used.

Common terminologies for reporting statistical tests (e.g with StatO) enable reuse and reproducibility.

Repositories

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

1. For human gene expression:
    1.1. Transcriptomics: EGA European Genome-Phenome Archive (if the data must 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 other relevant repositories can be found in FAIRsharing using the query' 'transcriptomics'.
    1.2. Gene expression arrays: Expression Atlas; GEO; All Of gene Expression; ArrayExpress
2. Cell lines/Animals: ArrayExpress/European Nucleotide Archive (ENA), Japanese Genotype-phenotype Archive,/DDBJ Sequence Read Archive or NCBI GEO/SRA
3. Genome-wide association studies (GWAS): GWAS Catalog; EGA; GWAS Central
4. Adaptive Immune Receptor Repertoire sequencing (AIRR-seq) data and annotations: AIRR Data Commons; NCBI

Data and metadata standards

1. Gene expression
    1.1. Transcriptomics.
        1.1.1. Preferred minimal metadata standard MINSEQE
        1.1.2. Preferred file formats (sequencing-based):
            1.1.2.1. Raw sequences: fastq (compression can be added with gzip)
            1.1.2.2. Mapped sequences: .sam (compression with .bam or .cram)
            1.1.2.3. Transcript count: TPM .csv
        1.1.3. Also see FAIRsharing using the query 'transcriptomics'
    1.2. Microarrays:
        1.2.1. Preferred minimal metadata standard: MIAME.
        1.2.2. Preferred file formats tab-delimited text, raw data file formats from commercial microarray platforms (Affymetrix, Illumina etc)
2. GWAS
    2.1. Preferred minimal metadata standard: MIxS
    2.2. Preferred file formats: for binary files: .bim .fam and .bed; for text-format files .ped and .map.
3. Adaptive Immune Receptor Repertoire sequencing (AIRR-seq)[1].
    3.1. Preferred minimal metadata standards: MiAIRR
    3.2. Preferred file formats:
        3.2.1. AIRR repertoire metadata (formatted as .JSON or .YAML), AIRR rearrangements (formatted as .TSV)
    3.3. Also see FAIRsharing using the query 'AIRR'.

Recommendations for policy makers

1. Although there is a growing number of consortia for genetic determinants (See https://www.covid19hg.org/), due to the high costs involved with advanced high-throughput genomics, data is disproportionately not available from Low and Middle Income Countries (LMICs) thus leading to improper extrapolation of results to population of unrepresented population groups. A strong policy framework is required to facilitate research and encourage participation from LMICs.

[1] Adaptive Immune Receptor Repertoire sequencing (AIRR-seq) samples the diversity of the immunoglobulins/antibodies and T cell receptors present in a host. The respective gene loci undergo random and irreversible rearrangement during lymphocyte development, therefore this data is fundamentally distinct from conventional genome sequencing.

7.2.3 Recommendations for Structural data

Repositories

1. Structural data on proteins acquired using any experimental technique is stored and found in the wwPDB: Worldwide Protein Data Bank; a collaborating cluster of three regional centers at (1) Europe EBI PDBe (PDBe-KB) and The Electron Microscopy Data Bank EMDB; (2) USA RCSB PDB; and (3) Japan PDBj. Data submitted to either of these resources will be available through each of them.

Data and metadata standards

1. X-ray diffraction
   1.1. There are no widely accepted standards for experimental raw data files. Generally these are stored and archived in the Vendor's native formats. Metadata is stored in CBF/imgCIF format (See: catalogue of metadata resources for crystallographic applications)
   1.2. Processed structural information is submitted to structural databases in the .pdf or .mmCIF format.
2. Electron microscopy
   2.1. Processed structural information is submitted to structural resources as PDBx/mmCIF
3. Molecular Dynamics (MD)
   3.1. Raw trajectory files containing all the coordinates, velocities, forces and energies of the simulation are stored as binary files: .trr, .dcd, and .netCDF ; See also a description of metadata standards to be considered.

7.3 Additional Working Documents & Links

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

7.4 References

AIRR Common Repository Working Group (CRWG) AIRR Data Commons API V1 — AIRR Standards 1.3.0 documentation. AIRR Data Commons API V1 — AIRR Standards 1.3.0 documentation, https://docs.airr-community.org/en/latest/api/adc_api.html

AIRR Minimal Standards Working Group (2020) AIRR Community Metadata Standard. https://github.com/airr-community/airr-standards

Artic Network Artic Network: hCoV-2019 (nCoV-2019/SARS-CoV-2). Artic Network, https://artic.network/ncov-2019

BBMRI-NL initiative Integrative Omics data set | BBMRI. Integrative Omics data set | BBMRI, https://bbmri.nl/services/samples-images-data/integrative-omics-data-set

Berners-Lee T (2009) Linked Data - Design Issues. Linked Data - Design Issues, https://www.w3.org/DesignIssues/LinkedData.html

Biocomputing Unit - Centro Nacional de Biotecnología (CNB), 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

COVID-19 Biohackathon April 5-11 Participants, University of Manchester, HITS gGmbH A COVID-19-specific instance for EOSC-Life's WorkflowHub. The WorkflowHub, https://covid19.workflowhub.eu/

COVID19-hg (2020) COVID-19 host genetics initiative. https://www.covid19hg.org/

DDBJ Center Bioinformation and DDBJ Center. Bioinformation and DDBJ Center, https://www.ddbj.nig.ac.jp/index-e.html

Di Cosmo R, Gruenpeter M, Zacchiroli S (2018) Identifiers for Digital Objects: the Case of Software Source Code Preservation. iPres2018, :9.
DNAstack COVID-19 Beacon. COVID-19 Beacon, https://covid-19.dnastack.com/_/discovery?position=3840&referenceBases=A&alternateBases=G
Dutch Techcentre for Life Sciences Personal Health Train. Dutch Techcentre for Life Sciences, https://www.dtls.nl/fair-data/personal-health-train/
EGA European Genome-Phenome Archive. The European Genome-phenome Archive, https://ega-archive.org/
ELIXIR COVID-19: The bio.tools COVID-19 Coronavirus tools list. bio.tools · Bioinformatics Tools and Services Discovery Portal, https://bio.tools/t?domain=covid-19
European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) EMBL-EBI COVID-19 Data Portal. https://www.covid19dataportal.org/
European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) EMBL-EBI COVID-19 Data Portal Introduction. https://www.ebi.ac.uk/covid-19
European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) European Nucleotide Archive. European Nucleotide Archive < EMBL-EBI, https://www.ebi.ac.uk/ena
European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) Expression Atlas: Gene Expression across species and biological conditions. Home < Expression Atlas < EMBL-EBI, https://www.ebi.ac.uk/gxa/home
European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) Pathogens: Surveillance, Identification Investigation. Pathogens, https://www.ebi.ac.uk/ena/pathogens/covid-19
European Nucleotide Archive How to Submit Raw Reads — ENA Training Modules 1 documentation. https://ena-docs.readthedocs.io/en/latest/submit/reads.html
European Nucleotide Archive SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Submissions — ena-browser-docs latest documentation. https://ena-browser-docs.readthedocs.io/en/latest/help_and_guides/sars-cov-2-submissions.html
FAIRsharing FAIRsharing. Oxford e-Research Centre University of Oxford, https://fairsharing.org/
FAIRsharing Team (2020) iReceptor Public Archive. FAIRsharing, https://doi.org/10.25504/FAIRSHARING.EKDQE5
Freunde von GISAID e.V. ("GISAID") GISAID: Genomic epidemiology of hCoV-19. GISAID - Next hCoV-19 App, https://www.gisaid.org/epiflu-applications/next-hcov-19-app/
Galaxy team COVID-19 analysis on usegalaxy. COVID-19 analysis on usegalaxy, https://covid19.galaxyproject.org/
GenBank (2020) SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. U.S. Center for Disease Control, https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/
Genomics Standards Consortium (2019) Minimum Information about any (X) Sequence (MIxS) specification. Genomics Standards Consortium / mixs, https://github.com/GenomicsStandardsConsortium/mixs
GitHub Choose an open source license. Choose a License, https://choosealicense.com/
GitHub, Inc. (2016) Making Your Code Citable. Making Your Code Citable; GitHub Guides, https://guides.github.com/activities/citable-code/
Global Alliance for Genomics and Health GA4GH: Data Security Toolkit. Data Security Toolkit, https://www.ga4gh.org/genomic-data-toolkit/data-security-toolkit/
Global Alliance for Genomics and Health GA4GH: Genomic Data Toolkit. Genomic Data Toolkit, https://www.ga4gh.org/genomic-data-toolkit/
Global Alliance for Genomics and Health GA4GH: Regulatory & Ethics Toolkit. Regulatory & Ethics Toolkit, https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/

Global Alliance for Genomics and Health Global Alliance for Genomics and Health (GA4GH). https://www.ga4gh.org/

Hatcher EL, Zhdanov SA, Bao Y, Blinkova O, Nawrocki EP, Ostapchuck Y, Schäffer AA, Brister JR (2017) Virus Variation Resource - improved response to emergent viral outbreaks. Nucleic Acids Research, 45(D1):D482–D490. https://doi.org/10.1093/nar/gkw1065

IRMACS Centre at Simon Fraser University iReceptor. Log In | iReceptor, https://gateway.ireceptor.org/login

Jiménez RC, Kuzak M, Alhamdoosh M, Barker M, Batut B, Borg M, Capella-Gutierrez S, Chue Hong N, Cook M, Corpas M, Flannery M, Garcia L, Gelpí JLl, Gladman S, Goble C, González Ferreiro M, Gonzalez-Beltran A, Griffin PC, Grüning B, Hagberg J, Holub P, Hooft R, Ison J, Katz DS, Leskošek B, López Gómez F, Oliveira LJ, Mellor D, Mosbergen R, Mulder N, Perez-Riverol Y, Pergl R, Pichler H, Pope B, Sanz F, Schneider MV, Stodden V, Suchecki R, Svobodová Vařeková R, Talvik H-A, Todorov I, Treloar A, Tyagi S, Gompel M van, Vaughan D, Via A, Wang X, Watson-Haigh NS, Crouch S (2017) Four simple recommendations to encourage best practices in research software. F1000Research, 6https://doi.org/10.12688/f1000research.11407.1

Kovalsky A (2020) COVID-19 workspaces, data and tools in Terra. COVID-19 workspaces, data and tools in Terra - Terra Support, http://support.terra.bio/hc/en-us/articles/360041068771

Lamprecht A-L, Garcia L, Kuzak M, Martinez C, Arcila R, Martin Del Pico E, Dominguez Del Angel V, Sandt S van de, Ison J, Martinez PA, McQuilton P, Valencia A, Harrow J, Psomopoulos F, Gelpi JLl, Chue Hong N, Goble C, Capella-Gutierrez S (2019) Towards FAIR principles for research software. Data Science, :1–23. https://doi.org/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

Majovski R (2020) Broad scientists release COVID-19 best-practices workflows and analysis tools in Terra. Terra Support, http://support.terra.bio/hc/en-us/articles/360040613432

Martinez-Martin N, Magnus D (2019) Privacy and ethical challenges in next-generation sequencing. Expert Review of Precision Medicine and Drug Development, 4(2):95–104. https://doi.org/10.1080/23808993.2019.1599685

MPEG, the Moving Picture Experts Group., ISO/IEC JTC1/SC29/WG11 (2018) White paper on the objectives and benefits of the MPEG-G standard. https://mpeg.chiariglione.org/sites/default/files/files/standards/docs/w15047-v2-w15047_GenomeCompressionStorage.zip

National Center for Biotechnology Information (NCBI) NCBI Virus: Severe acute respiratory syndrome-related coronavirus, taxid:694009. NCBI Virus, https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome-related%20coronavirus,%20taxid:694009

National Center for Biotechnology Information (NCBI) NCBI Virus: Submit Sequences. Submit Sequences | NCBI Virus, https://www.ncbi.nlm.nih.gov/labs/virus/vssi/docs/submit/

National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) Submission Quick Start. SRA Submission Quick Start, https://www.ncbi.nlm.nih.gov/sra/docs/submit/

National Genomics Data Center 2019nCovR - China National Center for Bioinformation. https://bigd.big.ac.cn/ncov?lang=en

National Institute of Allergy and Infectious Disease (NIAID) (2013) Data Sharing and Release Guidelines. https://www.niaid.nih.gov/research/data-sharing-and-release-guidelines

Nextstrain Team, Bedford T, Neher R (2020) Nextstrain Genomic epidemiology of novel coronavirus. https://nextstrain.org/ncov/global

NIAID DAIT (2018) ImmPort Shared Data. ImmPort Shared Data, https://immport.org/shared/home
Nüst D, Sochat V, Marwick B, Eglen S, Head T, Hirst T (2020) Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science. https://doi.org/10.31219/osf.io/fsd7t
Opensource.org Licenses & Standards | Open Source Initiative. https://opensource.org/licenses
Piver LIPIDMAPS. http://www.lipidmaps.org
Rambaut A (2020) Phylogenetic analysis of nCoV-2019 genomes. http://virological.org/t/phylodynamic-analysis-176-genomes-6-mar-2020/356
Rambaut A Virological: Novel 2019 coronavirus discussion forum. Virological, http://virological.org/c/novel-2019-coronavirus
RCSB Protein Data Bank RCSB Protein Data Bank SARS-CoV-2 Resources. https://www.rcsb.org/news?year=2020&article=5e74d55d2d410731e9944f52&feature=true
RDA-COVID19-Omics Subgroup (2020) RDA-COVID19-Omics. RDA, https://www.rd-alliance.org/groups/rda-covid19-omics
Renieri A GEN-COVID: Impact of Host Genome on COVID-19 Clinical Variability. GEN-COVID, https://sites.google.com/dbm.unisi.it/gen-covid
Shanghai Public Health Clinical Center & School of Public Health (2020) Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. http://www.ncbi.nlm.nih.gov/nuccore/MN908947.3
SOLID Understanding Linked Data. https://solid.github.io/understanding-linked-data/#1
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
Technical Committee : ISO/TC 215/SC 1 Genomics Informatics ISO/TS 20428:2017 Health informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records. ISO, https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/79/67981.html
Technical Committee : ISO/TC 276 Biotechnology ISO/AWI 20688-2: Biotechnology — Nucleic acid synthesis — Part 2: General definitions and requirements for the production and quality control of synthesized gene fragment, gene, and genome. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/58/75852.html
The Broad Institute of MIT and Harvard Genotype-Tissue Expression (GTEx) Portal. GTEx Portal, https://www.gtexportal.org/home/
The Human Protein Atlas consortium SARS-CoV-2 related proteins - The Human Protein Atlas. https://www.proteinatlas.org/humanproteome/sars-cov-2
UniProt COVID-19 UniProtKB. UniProt, https://covid-19.uniprot.org/uniprotkb?query=*
Vander Heiden JA, Marquez S, Marthandan N, Bukhari SAC, Busse CE, Corrie B, Hershberg U, Kleinstein SH, Matsen IV FA, Ralph DK, Rosenfeld AM, Schramm CA, The AIRR Community, Christley S, Laserson U (2018) AIRR Community Standardized Representations for Annotated Immune Repertoires. Frontiers in Immunology, 9:2206. https://doi.org/10.3389/fimmu.2018.02206
wwPDB consortium, Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo LD, Christie C, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranovic V, Guzenko D, Hudson BP, Liang Y, Lowe R, Peisach E, Periskova I, Randle C, Rose A, Sekharan M, Shao C, Tao Y-P, Valasatava Y, Voigt M, Westbrook J, Young J, Zardecki C, Zhuravleva M, Kurisu G, Nakamura H, Kengaku Y, Cho H, Sato J, Kim JY, Ikegawa Y, Nakagawa A, Yamashita R, Kudou T, Bekker G-J, Suzuki H, Iwata T, Yokochi M, Kobayashi N, Fujiwara T, Velankar S, Kleywegt GJ, Anyango S, Armstrong DR, Berrisford JM, Conroy MJ, Dana JM, Deshpande M, Gane P, Gáborová R, Gupta D, Gutmanas A, Koča J, Mak L, Mir S, Mukhopadhyay A, Nadzirin N, Nair S, Patwardhan A, Paysan-Lafosse T, Pravda L, Salih O, Sehnal D, Varadi M, Vařeková R, Markley JL, Hoch JC, Romero PR, Baskaran K, Maziuk D,

Ulrich EL, Wedell JR, Yao H, Livny M, Ioannidis YE (2019) Protein Data Bank: the single global archive for 3D macromolecular structure data. Nucleic Acids Research, 47(D1):D520–D528. https://doi.org/10.1093/nar/gky949
ZB MED - Information Centre for Life Sciences ZB MED COVID-19 Hub. https://covid-19.zbmed.de/

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 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. This requires appropriate ethical and legal considerations. 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 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 for Researchers

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
        1.1.2. FAIR Software guidelines on code repositories
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
        2.1.2. List of software registries
        2.1.3. Making your code citable through GitHub and Zenodo
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
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.

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.
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
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

9.2.1 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/

9.3 Initial Guidelines for Policymakers

Research software is essential for research, and this is increasingly recognised globally by researchers. National and international policy changes are now needed to increase this recognition and to increase the impact of the software in important research and policy areas. This section provides recommendations for policy makers on how to support the research software community to respond to COVID-19 challenges, based on existing work (Akhmerov et al. 2019).

1. **Support the funding of development and maintenance of critical research software.** Policy makers must continue to allocate financial resources to programs that support the development of new research software and the maintenance of research software that has a large user base and/or an important role in a research area.
    1.1. Examples: UK Research and Innovation is funding COVID-19 related projects that can include work focussed on evaluation of clinical information and trials, spatial mapping and contact mapping tools (UK Research and Innovation 2020). Mozilla has created a COVID-19 Solutions Fund for open source technology projects (Mozilla 2020). USA's National Institutes of Health (NIH) provides "Administrative Supplements to Support Enhancement of Software Tools for Open Science" (NIH 2020b). The Chan Zuckerberg Initiative is funding open source software projects that are essential to biomedical research (Chan Zuckerberg Initiative 2020).
2. **Encourage research software to be open source and require it to be available.** Policy makers should enact policies that encourage provision of an open source license, or at least require it to be accessible. All research software should be released under a licence to ensure clarity of how it can be used and to protect the developers. The use of open source licences should be seen as the default for research software and policy makers should enact policies to encourage that practice. When software is made available under an open source license it means that its underlying code is made freely accessible , as encouraged by the "A" in FAIR (Findable, Accessible, Interoperable and Reusable) to users to examine and can be modified and redistributed. Through this process, software users can review, understand, improve, and build upon the software. As research outcomes rely on software, if software is not open source it must minimally be available for experimentation, to enable understanding of the software's functionality and properties and to reproduce the research outcomes. Whilst preprints and papers are increasingly openly shared to accelerate COVID-

19 responses, the software and code for these papers is often not cited and hard to find, making reproducibility of this research challenging, if not impossible (Smith et al. 2016). Encouraging publishers to make software availability a default condition, together with the usually existing requirement for data availability, is an excellent way to greatly improve this.

2.1. Examples: The research community has been increasing access to key software and code, such as the Imperial College epidemic simulation model that is being utilised by government decision-makers. This was made publicly available with support by Microsoft to accelerate the process (Adam 2020).

3. **Encourage the research community's ability to apply best practices for research software, including training in software development concepts**. Policy makers should provide programs and funding opportunities that encourage both researchers and research support professionals (such as Research Software Engineers and Data Stewards) to utilise best practices to develop better software faster. In order to make research software understandable and reusable, it must be produced and maintained using standard practices that follow standard concepts, which can be applied to software ranging from researchers writing small scripts and models, to teams developing large, widely-used platforms. As research is becoming data-driven and collaborative in all areas, all researchers would benefit from the development of core software expertise, and research support professionals with these expertise also need to be increased. Policy makers should support inclusive software skills and training programs, including development of communities of learners and trainers.

3.1. Examples: There are various initiatives that link community members with specific digital skills to projects needing additional support, including Open Source Software helpdesk for COVID-19 ("COVID-19 OSS HELP" 2020) and COVID-19 Cognitive City (Grape 2020). Other initiatives aim to increase skills for engaging with software and code, such as the Carpentries, USA's NIH events (NIH 2020a); and the Galaxy Community and ELIXIR's webinar series (ELIXIR 2020).

4. **Support recognition of the role of software in achieving research outcomes.** Policy makers should enact policies and programs that recognise the important role of research software in achieving research outcomes. It is important that policy makers encourage the development of research assessment systems that reward software outputs, alongside publications, data and other research outputs; and ensure that data and software management plans are a requirement in funding processes. It is also important that policy makers work to ensure these systems include proactive responses when these are not implemented.

4.1. Examples: Policy makers need to support initiatives such as the Declaration on Research Assessment (DORA n.d.), which are beginning to be utilised by research agencies including the Wellcome Trust (Wellcome 2020), signatories to the Concordat to Support the Career Development of Researchers (Vitae 2020).

9.3.1 References

Adam, David. 2020. "Special Report: The Simulations Driving the World's Response to COVID-19." *Nature* 580 (7803): 316–18.

https://doi.org/10.1038/d41586-020-01003-6.
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." https://doi.org/10.5281/zenodo.3378572.
Chan Zuckerberg Initiative. 2020. "CZI Launches Funding Opportunity for Open Source Software." *Chan Zuckerberg Initiative* (blog). April 30, 2020. https://chanzuckerberg.com/rfa/essential-open-source-software-for-science/.
"COVID-19 OSS HELP." 2020. April 30, 2020. https://covid-oss-help.org/.
DORA. n.d. "San Francisco Declaration on Research Assessment," n.d. https://sfdora.org/read/.
ELIXIR. 2020. "Galaxy-ELIXIR Webinar Series: FAIR Data and Open Infrastructures to Tackle the COVID-19 Pandemic | ELIXIR." April 30, 2020. https://elixir-europe.org/events/webinar-galaxy-elixir-covid19.
Grape, Derek. 2020. "Exaptive Partners with Bill & Melinda Gates Foundation, Launching The COVID-19 Cognitive City." April 30, 2020. https://www.exaptive.com/news/exaptive-partners-with-bill-melinda-gates-foundation-launching-the-covid-19-cognitive-city.
Mozilla. 2020. "MOSS Launches COVID-19 Solutions Fund." The Mozilla Blog. April 30, 2020. https://blog.mozilla.org/blog/2020/03/31/moss-launches-covid-19-solutions-fund.
NIH. 2020a. "NIH to Host Webinar on Sharing, Discovering, and Citing COVID-19 Data and Code in Generalist Repositories on April 24 | Data Science at NIH." April 30, 2020. https://datascience.nih.gov/news/nih-to-host-webinar-on-sharing-discovering-and-citing-covid-19-data-and-code-in-generalist-repositories-on-april-24.
———. 2020b. "NOT-OD-20-073: Notice of Special Interest (NOSI): Administrative Supplements to Support Enhancement of Software Tools for Open Science." April 30, 2020. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-073.html.
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.
UK Research and Innovation. 2020. "Get Funding for Ideas That Address COVID-19." April 30, 2020. https://www.ukri.org/funding/funding-opportunities/ukri-open-call-for-research-and-innovation-ideas-to-address-covid-19/.
Vitae. 2020. "Concordat to Support the Career Development of Researchers — Vitae Website." Landing page. April 30, 2020. https://www.vitae.ac.uk/policy/concordat-to-support-the-career-development-of-researchers.
Wellcome. 2020. "Outputs Management Plan - Grant Funding | Wellcome." April 30, 2020. https://wellcome.ac.uk/grant-funding/guidance/how-complete-outputs-management-plan.

9.4 Additional Working Documents & Links

Additional materials from the Working Group available: RDA-COVID19-Software

10. Overarching Legal and Ethical Guidelines

The intention of these guidelines is to help researchers, practitioners and policy-makers deal ethically and within legal constraints with all aspects of pandemic response and in particular with regard to key ethical principles of *equity, utility, efficiency, liberty, reciprocity and solidarity*[1]. In times of public health emergency, it is right to consider how best to respond in terms of increased data and research outcome sharing. However, this does not mean that the principle of law[2] may be overlooked, nor that overarching concerns related to human rights and dignity[3], especially where this may lead to marginalisation or other forms of stigmatisation, should not inform appropriate research principles directed towards the common good.

These guidelines have been produced by the RDA-COVID working group in March-April 2020 during the ongoing coronavirus disease (COVID-19) pandemic. The aim is to identify and collate existing recommendations and guidelines in order to increase the speed of scientific discovery by enabling researchers and practitioners to:

1. Readily identify the guidance and resources they need to support their research work
2. Understand generic and cross-cutting ethical and legal considerations
3. Appreciate country- or region-specific differences in policy or legal instruments
4. Identify the institutional stakeholders best placed to provide relevant ethical and legal guidance

10.1 Sub-Group Focus and Description

The COVID-19 pandemic has created significant confusion for researchers in terms of whether, and in which way, existing Ethical, Legal and Social Implication (ELSI) principles remain relevant. In principle, the COVID pandemic does not constitute an emergency in relation to which all existing ELSI documents and principles lose validity. The COVID pandemic does not serve to remove the basic validity of the rights and interests on which these documents and principles are based. The emergency does, however, mandate a reconsideration of the balance between these rights and interests - in particular between research subject's right to privacy and the public interest in the outcome of research. In some cases, this reconsideration has led to legitimate adaptations of, or derogation from, normally applicable principles.

10.2 Highlights

This section provides a high-level overview of:

1. Cross cutting principles
2. Hierarchy of norms
3. Where to seek guidance
4. Existing relevant policy statements

Cross-cutting principles

All activities, especially in times of pandemic or other public emergencies, should be guided by:

1. The FAIR (Findable, Accessible, Interoperable and Re-usable) principles of data to ensure ongoing, beneficial research[4];
2. The CARE principles to ensure ethical treatment of individuals and communities[5];
3. The Global Code of Conduct, specifically Fairness, Respect, Care and Honesty in research activities, to maximise equanimity in research outcome benefit[6].

Hierarchy of obligations

ELSI considerations related to research are elaborated in three key types of document: ethical guidelines; policy guidance; and legal instruments. The distinction between these types of instrument is not always obvious. The following principles are useful for COVID-19 researchers considering the interaction between instruments:

1. Ethical guidelines are often defined and publicised by non-law-making bodies, while legal instruments will be adopted by governments or other legislative bodies.
2. Many ethical instruments are mandatory for researchers or clinicians, such as those imposed by professional orders, healthcare institutions, or governmental funding agencies.
3. Instruments exist in a hierarchy, with legal instruments assumed to take precedence. It should be remembered, however, that it is usually possible to continue to work ethically whilst adhering to the law.
4. Jurisprudence and other official guidelines providing authoritative interpretations of legal instruments will also take precedence over ethical guidelines.
5. When legal instruments do not address a point, or leave room for interpretation on a point, ethical instruments can be used to guide practice.

Common obligations in using health data that are found in many laws and ethical guidelines include the following:

1. The obligation to preserve confidentiality
2. The obligation to ensure data accuracy
3. The obligation to use anonymised data instead of personal data, or minimise personal data use
4. The obligation to limit the identifiability of personal data as far as possible - including via pseudonymisation techniques
5. The need to process for a specific, authorized, purpose and to only process for secondary purposes provided certain conditions are fulfilled
6. To hold oneself accountable to, and remain transparent towards, the individuals concerned by the data used
7. To provide individuals access to their data, and to rectify errors or biases in the data on request
8. To provide individuals the opportunity to request the deletion or return of their data in certain circumstances

Researcher ethical obligations may be summarised as:

1. To ensure non-malevolence
2. To aim to maximise benevolence
3. To ensure fair, balanced and timely access to research outcomes
4. Wherever possible, to respect individual autonomy
5. To share data and research findings
6. To support cross-disciplinary collaboration

Such obligations are formalised through ethical guidance[7]. Especially in times of pandemic specific attention to vulnerable groups and guidance on related global justice issues are to be commanded.

Seeking guidance

In times of pandemic or other public emergencies, it is important to be aware of existing and *ad hoc* resources and guidance. For example:

1. For researchers attached to an academic institution:
   1.1. the Institutional Review Board (IRB) or Research Ethics Committee (REC) will provide guidance as well as review;
   1.2. the Information Governance Board will provide support on data management;
   1.3. the Data Protection Officer will provide support and guidance on data protection issues;
   1.4. technology transfer offices provide guidance regarding intellectual property and related issues.
2. For professionals affiliated to a professional body, the latter will provide guidance on ethical research activities.
3. For medical or other clinical staff, the institution (such as a hospital) will provide research integrity support, including ethical approvals required and an ad hoc mechanisms to support emergency research efforts; or the appropriate governing body (e.g., the NHS in the UK) will provide training and support both ongoing and in exceptional circumstances.
4. Hospitals, much like academic institutions, are often staffed by a Data Protection Officer, personnel specialized in research ethics including IRBs or REBs, and administrators responsible for authorizing the sharing of health data.

Researchers and other professionals should always consult their institutional support personnel as well as professional bodies.

Relevant policy and non-policy statements

The RDA Covid-19 Ethical-Legal group endorses and recommends guidance published as follows:

1. the OECD Privacy Principles
2. the UNESCO World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) in their STATEMENT ON COVID-19: ETHICAL CONSIDERATIONS FROM A GLOBAL PERSPECTIVE
3. the Council of Europe pointers to national resources from national ethics committees or other related to COVID-19
4. the Council of Europe statement on bioethics during COVID-19
5. the European Group on Ethics in Science and New Technologies statement on solidarity
6. the Global Alliance for Genomics and Health (GA4GH) Framework for Responsible Sharing of Genomic and Health-Related Data

10.3 Initial Sub-Group Guidelines

Work in progress

The ELSI group are currently expanding upon initial discussion and investigations to develop the following:

Specific

1. What do we already know to be part of *best practice*?
2. What are the common pitfalls that we are aware of?

Aspirational

1. What would we like to see?
2. What should be looked at in the future?

Data collection principles and practice

1. Consent
2. Provenance
3. Trustworthiness
4. Community vs Individual

Data sharing

1. Cross-border sharing
2. Equitable sharing
3. Cross-disciplinary sharing

Emergency provisions

1. What is available during emergencies
2. What must still be done / avoided
3. How to future-proof data and findings

Resources

1. What is available
2. How to find it
3. Who to ask?

The WG welcomes requests from other groups and disciplines about specific issues and concerns they may encounter in their own domains.

Additional materials from the Working Group available: RDA-COVID19-Legal-Ethical

10.4 References

[1] https://apps.who.int/iris/bitstream/handle/10665/70006/WHO_CDS_EPR_GIP_2007.2_eng.pdf
[2] https://edpb.europa.eu/our-work-tools/our-documents/other/statement-processing-personal-data-context-covid-19-outbreak_en
[3] https://www.coe.int/en/web/human-rights-rule-of-law/covid19
[4] https://www.force11.org/fairprinciples
[5] http://www.newcarestandards.scot/?page_id=15
[6] https://www.globalcodeofconduct.org/
[7] For example, the Caldicott principles (https://www.igt.hscic.gov.uk/Caldicott2Principles.aspx), Medical Ethics (Beauchamp & Childress (2001) Principles of Biomedical Ethics; ISBN 9780195143324)