RDA COVID-19 Working Group

Recommendations and Guidelines
4th Release
15 May 2020

RDA Recommendation (4th Release - 15 May 2020)

Produced by: **RDA COVID-19 Working Group, 2020**

Document Metadata

Identifier DOI: 10.15497/RDA00046

Title RDA COVID-19; recommendations and guidelines, 4th release 15 May 2020

Description This is the fourth 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 and Legal and Ethical .

Date Issued 2020-05-15

Version Draft guidelines and recommendations; fourth release, 15 May 2020, version for public review

Contributors RDA COVID-19 Working Group
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.

Disclaimer The views and opinions expressed in this document are those of the individuals identified (below), and do not necessarily reflect the official policy or position of their respective employers, or of any government agency or organization.

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, Christian Ohmann
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: *Christophe 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 subgroup documents.

The views and opinions expressed in this document are those of the individuals identified (below), and do not necessarily reflect the official policy or position of their respective employers, or of any government agency or organization.

Table of Contents

Table of Figures 6

1. Objectives and Use of This Document 8

2. Foundational Recommendations 10

2.1 Challenges 10

2.2 Recommendations 11

3. Data Sharing in Clinical Medicine 13

3.1 Focus and Description 13

3.2 Scope 13

3.3 Policy Recommendations 13

3.3.1 Sharing Clinical Data 13

3.3.2 Trustworthy Sources of Clinical Data 14

3.4 Guidelines 17

3.4.1 Clinical Trials on COVID-19 17

3.4.2 Other Types of Data 18

4. Community Participation and Data Sharing 20

4.1 Focus and Description 20

4.2 Scope 21

4.3 Policy Recommendations 22

4.3.1 Data Collection 22

4.3.2 Data Quality and Documentation 23

4.3.3 Legal and Ethical Aspects 23

4.3.4 Software Development 23

4.4 Guidelines 23

4.4.1 Data Collection 24

4.4.2 Data Quality and Documentation 24

4.4.3 Data Storage and Long-Term Preservation 24

4.4.4 Transparency and Community Participation 25

5. Data Sharing in Epidemiology 26

5.1 Focus and Description 26

5.2 Scope 26

5.3 Recommendations 28

5.3.1 Policy 28

5.3.2 Information Technology And Data Management Infrastructure 28

5.3.3 Analysis and Modeling 28
5.3.4 Surveillance Data 28
5.3.5 Global preparation, detection, and response 28
5.3.6 Interoperability and Data Exchange 28
5.4 Guidelines 29
5.4.1 Policy 29
5.4.2 Information Technology and Data Management Infrastructure 29
5.4.3 Analysis and Modeling 30
5.4.4 Surveillance Data 30
5.4.6 Interoperability and Data Exchange 34
5.4.6.1 COVID-19 Questionnaire Initiatives 34
5.4.6.2 COVID-19 Question Bank 40
5.4.6.3 COVID-19 Use Case for Common Data Models 42
5.4.6.4 Anonymization and Pseudonymization 44
6. Data Sharing in Omics Practices 47
6.1 Focus and Description 47
6.2 Scope 47
6.3 Policy Recommendations 47
6.3.1 Researchers Producing Data 47
6.3.2 Funders 48
6.3.3 Publishers 49
6.3.4 Policymakers 49
6.3.5 Researchers 49
6.3.6 Providers of Data Sharing Infrastructures 49
6.4 Guidelines 50
6.4.1 Recommendations for Virus Genomics Data 50
6.4.1.1 Repositories 50
6.4.1.2 Data and Metadata Standards 50
6.4.2 Recommendations for Host Genomics Data 51
6.4.2.1 Generic Recommendations 52
6.4.2.2 Repositories 52
6.4.2.3 Data and Metadata Standards 53
6.4.3 Recommendations for Structural Data 54
6.4.3.1 Repositories 54

6.4.3.2 Locating Existing Data 54
6.4.3.3 Data and Metadata Standards 54
6.4.4 Recommendations for Proteomics 56
6.4.4.1 Repositories 56
6.4.4.2 Data and Metadata Standards 57
6.4.5 Recommendations for Metabolomics 57
6.4.5.1 Repositories 57
6.4.5.2 Data and Metadata Standards 57
6.4.6 Recommendations for Lipidomics 58
6.4.6.1 Generic Recommendations for Researchers 58
6.4.6.2 Repositories 58
6.4.6.3 Data and Metadata Standards 58
7. Data Sharing in Social Sciences 60
7.1 Focus and Description 60
7.2 Scope 60
7.3 Policy Recommendations 61
7.4 Guidelines 62
7.4.1 Data Management Responsibilities and Resources 62
7.4.2 Documentation, Standards, and Data Quality 63
7.4.3 Storage and Backup 64
7.4.4 Legal and Ethical Requirements 64
7.4.5 Data Sharing and Long-term Preservation 65
8. Research Software and Data Sharing 67
8.1 Focus and Description 67
8.2 Scope 67
8.3 Policy Recommendations 67
8.4 Guidelines for Publishers 70
8.5 Guidelines for Researchers 71
9. Legal and Ethical Compliance 73
9.1 Focus and Description 73
9.2 Scope 73
9.3 Policy Recommendations 74
9.3.1 Initial Recommendations 74
9.3.2 Relevant Policy and Non-Policy Statements 75

9.4 Guidelines 75
9.4.1 Cross-Cutting Principles 75
9.4.2 Hierarchy of Obligations 76
9.4.3 Seeking Guidance 77
9.4.4 Anonymisation 78
9.4.5 Consent 80
9.4.6 The 5 Safes Model: Safe People, Safe Projects, Safe Settings, Safe Outputs, Safe Data 81
10. Additional Working Documents 83
11. References 86
12. Contributors 107

Table of Figures

Figure 1. 8
Figure 2. 27
Figure 3. 32
Figure 4. 34
Figure 5. 41
Figure 6. 45
Figure 7. 45
Figure 8. 46

Table of Tables

Table 1 Questionnaire instruments: Reference studies. Development of such initiatives is a very active and rapidly changing area at the present time during the COVID-9 pandemic. Contributed by Dr. Carsten Schmidt, Dr Jay Greenfield, Dr. Stefen Sauermann, and Dr. Chifundo Kanjala, developed from discussions held during RDA-COVID19-Epidemiology Work Group meetings .35
Table 2 Questionnaire instruments - Resources. Development of such initiatives is a very active and rapidly changing area at the present time during the COVID-9 pandemic. 40

0. Log Changes

Document	Changes	Date
RDA COVID-19; recommendations and guidelines, 1st release 24 April 2020	First draft of document released for comments and feedback	24 April 2020
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 May 2020
RDA COVID-19; recommendations and guidelines, 3rd release 8 May 2020	Section 3 - Foundational Principles/Recommendations - updates and additional content Section 4 - Clinical - Revisions to sections 4.2.3, 4.2.4, and 4.3. Newly added section 4.2.5 Section 5 - Community Participation: Use case section 5.2.2 added Section 6 - Epidemiology - major revisions/additional content Section 7 - Omics - updated Section 8 - Social Sciences - revised Section 9 - Software - updated and sections added/reorganized (9.4, 9.5) Section 10 - Legal/Ethical - major revisions Incorporation of feedback received via Requests for Comments (RfC) process and directly to Co-chairs, moderators and editorial team	8 May 2020
RDA COVID-19; recommendations and guidelines, 4th release 15 May 2020	Overarching structural changes throughout the document Sections 1 & 2 - revisions Section 3 - Clinical - General edits and section restructuring in 3.3.1 & 3.3.2 Section 4- Community Participation - no changes after larger restructure Section 5 - Epidemiology - updates to all sections Section 6 - Omics - added section 6.2, revisions to 6.1, 6.3, 6.4 Section 7 - Social Sciences - added sections 7.2, 7.3; updates to 7.4 Section 8 - Software - Added to Section 8.2 Scope, Swapped 8.4 and 8.5, general text revisions throughout Section 9 - Legal/Ethical - added sections 9.4.3, 9.4.4, and 9.4.6 Section 10 – Additional working documents – new section added Section 11 - New Reference section added; includes references tagged "*CITEDresource" in Zotero library Section 12 - Contributors - new section added	15 May 2020

1. Objectives and Use of This Document

During a pandemic, data combined with the right context and meaning can be transformed into knowledge for informing public health responses. 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 recommendations 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.

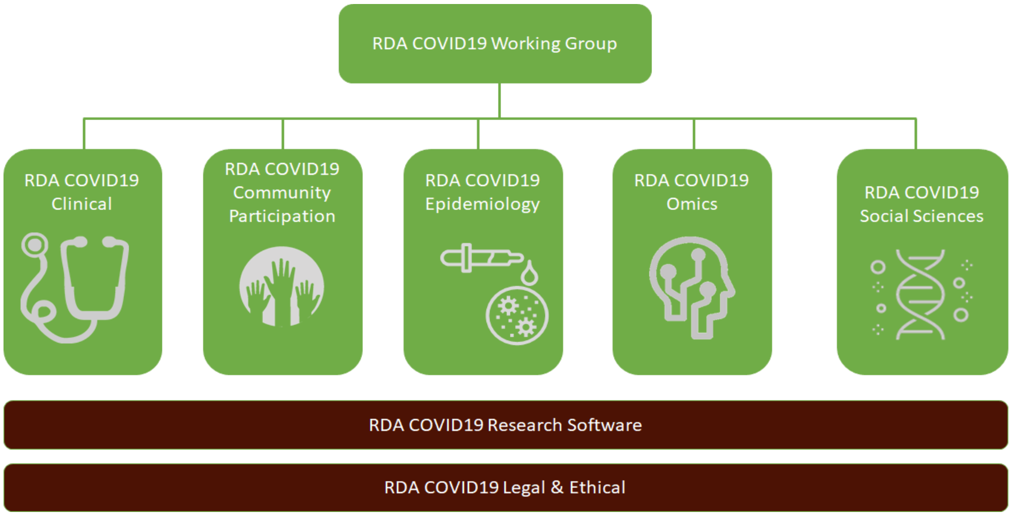
RDA COVID19 Working Group
RDA COVID19 Clinical
RDA COVID19 Community Participation
RDA COVID19 Epidemiology
RDA COVID19 Omics
RDA COVID19 Social Sciences
RDA COVID19 Research Software
RDA COVID19 Legal & Ethical

Figure 1. Research Data Alliance COVID-19 Working Group sub-groups including research areas and cross-cutting themes.

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 contained herein address general aspects related to the principles that data should adhere to, for example FAIR and the adoption of community standards, while also providing a tool which could help researchers and data stewards to determine the standards for what is ‘good enough’ when there is significant value to sharing research outputs as quickly as possible. The work has been divided into 5 research areas with two cross cutting themes, as a way to focus the conversations, and provide an initial set of guidelines in a tight timeframe.

The RDA COVID-19 WG was initiated after a conversation between the RDA Secretary General and the European Commission. 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 research sub-groups, and the first group meetings started taking place soon after, with cross-cutting themes quickly emerging.

As of 15 May, there are over 440 members of the CWG, relatively evenly spread across the 7 groups. The first three drafts were released and simultaneously opened for comment on 24 April, 1 May and 8 May, indicating huge progress in the space of 5 weeks. This 15 May version marks the release of a newly structured document, that attempts to retain the semantic uniqueness required by each research area, while also harmonising the format for ease of navigation. The fifth and final draft release will carry this integration further, attempting to identify possible elements that are common across most groups. The current timeline will see the 5th (final) release on 28 May, followed by a 10 day period for community feedback. The final endorsed release of Version 1 of the Recommended Guidelines is slated for 30 June.

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 and improve our work and will be incorporated into the sub-group discussions, and the next set of writing sprints.

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

2. Foundational Recommendations

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

2.1 Challenges

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 rapid pace of the disease and the immense and rapid mobilisation of resources could create an environment for inaccurate or low quality data, which could have considerable implications. For example:

1. shortcuts with the interpretation of data can create issues, such as the early debate on whether COVID-19 is 'just another flu' or not;
2. the 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, a lack of transparency and publication of false or unchecked numbers is perhaps worse than no publication at all.

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.

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).

2.2 Recommendations

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 effort should be put into the quality and documentation of the dataset.

Metadata

The key to finding and using digital assets is metadata. COVID-19 research requires access to different assets for different communities. Within a given community, the commonly used metadata standards are well-known, but a researcher working across communities has more difficulty in locating relevant assets. In this case a 'metadata element set' that is generally applicable is required to be associated with each asset so that they can be used under the FAIR principles. A proposed metadata element set is available on the RDA Metadata Interest Group page. At present there are four generic metadata standards that are used widely, Dublin Core (DC), DCAT, DataCite and Schema.org. The latter has a specialisation called Bioschemas which provides a way to add semantic markup to web pages for improved findability of data in the life sciences, and is currently updating profiles to aid in discovery of COVID-19 data. Providing FAIR access to assets would be much enhanced now if assets had metadata encoded in one of these standards – as well as in the metadata standard(s) used by the particular community. It is to be hoped that in future richer generic metadata standards will be used. For a longer registry of metadata standards, see the Metadata Standards Directory or the FAIRSharing 'Standards' section

However, the use of these standards for machine-to-machine communication depends on how they are implemented. Many DC implementations are in text, HTML or XML form and used more easily by human readability than machine understandability. More recent implementations use Resource Description Framework (RDF) which does provide machine-to-machine capability. Earlier DCAT implementations used XML, more recent implementations use RDF. DataCite uses XML but also schema.org metadata format and JSON-LD while Schema.org uses RDF and JSON-LD. Thus, these metadata standards encourage machine-to-machine interoperation.

Metadata has two aspects: syntax and semantics. The syntax defines the structure of the metadata information and should conform to a formal grammar. The semantics defines the meaning of strings of characters – usually through an associated ontology – and should be declared. Again, there are generic ontologies (or vocabularies which have less detail on relationships between the terms) and community-specific ontologies (or vocabularies).

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. The recent joint statement on the Duty to Document underlines how crucial it is, especially during this time of rapid and unprecedented decision making, to document decisions, and secure and preserve records and data for the future.

Use of Trustworthy Data 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 or FAIRSharing 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 are listed here.

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.

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.

3. Data Sharing in Clinical Medicine

3.1 Focus and Description

Clinical activities are at the forefront of combating 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 centered on obtaining consent to address future use of data, conducting clinical trials, sharing clinical information (personal and health data) and ensuring that results are shared and reused in a trustworthy and efficient manner. For this 3rd release, recommendations for consent and clinical trials have been slightly modified, recommendations for clinical data sharing have been extensively revised and some elements for immunological data and imaging data have been added and will be further supplemented.

3.2 Scope

Access to clinical trial information and clinical data is critical to accelerating research and response to the novel coronavirus (COVID-19). Collecting data on individuals presenting with suspected or confirmed COVID-19 is essential to improved patient care and informing the public health response. Data collection and procedures on data sharing specific for COVID-19 should be standardized to identify key risk factors and focus on specific risk factors in accordance with commonly accepted data standards. Specifically, consent for clinical trials should be obtained in accordance with ISO/TS 17975:2015 [1] (Health informatics — Principles and data requirements for consent in the Collection, Use or Disclosure of personal health information) and corresponding national legislations.

Sharing genomic and health-related data should follow recommendations modeled after the Global Alliance for Genomics and Health (GA4GH) Consent Policy. Access to sensitive personal data (e.g.genetic data, health- related data) should be outlined in Data Access Agreements (DAAs) between the data holder and secondary data users, and data requests should be reviewed and managed by Data Access Committees to determine whether future data uses are consistent with data use limitations. In addition, data should be shared in accordance with applicable laws, regulations, and policies. More information on the ethical and legal bases will be found in the chapter from the legal and ethical sub-group.

3.3 Policy Recommendations

3.3.1 Sharing Clinical Data

General Aspects

In the COVID-19 situation, promotion of data sharing is of utmost importance because many studies are performed under enormous time pressure, with weaknesses in the methodology (e.g. no control) and preliminary results published without any review. Sharing of data, and related documentation (e.g. protocols) will reduce duplication of effort and improve trial design, when many similar studies are being planned or implemented in different countries. Clinical data outside clinical trials (e.g., case studies, descriptive cohorts of patients, etc.) may also be of high value and should be reported using appropriate reporting guidelines (see EQUATOR Network guidelines and FAIR Sharing Registry).

3.3.2 Trustworthy Sources of Clinical Data

During a pandemic like COVID-19, it is important to concentrate efforts on scrutinizing reliable data sources that provide data and metadata of high quality and guarantee the authenticity and integrity of the information. The recommendations are:

1. Data and trial documents should be transferred to a suitable and secure data repository to help 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.
2. 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[7] [8]or accreditation. A wide -range of community-based standardized quality criteria, best practices, and principles (e.g. TRUST Principles[2]) should also be considered.
3. If analysis environments that allow in situ analysis of data sets but prevent downloads are available, they should be provided to the end-user researchers, in a pandemic situation, without fees if possible.
4. Tools allowing different data sets from different repositories to be analysed together on a temporary basis should be provided.
5. Adequate tools should be implemented for collection and analysis of reliable real-world data on drugs approved for the treatment of COVID-19.

Data Standards

To maximize the value of clinical data and information, it is optimal to use consistent structuring to develop document and make data comply to FAIR principles, e.g. Findable, Accessible, Interoperable and Reusable. This is achieved by applying agreed domain-specific standards for data formats and semantic content (e.g. terminologies). Hence, these data and metadata standards support the consistent access to and reliable exchange of data from COVID-19 clinical research and case reporting.

1. Widely accepted data and metadata standards should be applied in COVID-19 studies and case reporting. Among the various standards for consistently defining, coding and reporting data from clinical research and case reports preferably those from the Clinical Data Interchange Standards Consortium (CDISC) and, especially for exchanging electronic health records (EHR), HL7 FHIR (Fast Healthcare Interoperability Resources) are encouraged to consider for ensuring data interoperability. Clinical trials, case

reports and public health studies should put the CDISC Interim User Guide for COVID-19 into consideration. For computational tools used in the clinical research and case reporting the application of COVID-19 specific FHIR profiles are recommended, if available. For cases where CDISC and HL7 standards are not applicable or feasible, there are alternatives, especially for academic teams. A comprehensive list of standards to format and describe clinical data and metadata is available at fairsharing.

2. Standardized clinical terminologies and ontologies should be used to describe the semantic content of the data and corresponding metadata, e.g. International Classification of Diseases (ICD) of WHO, SNOMED CT and LOINC (Logical Observation Identifiers Names and Codes). This ensures unambiguous interpretation (by humans and by computer algorithms) of the used terms describing the data and its elements. SNOMED CT and ICD-10 both were extended by specific terms corresponding to COVID-19 and special use codes are developed for LOINC that can be accessed as prerelease terms. Additional specific clinical terminologies and ontologies can be found at fairsharing.
3. More support is likely to be needed for academic researchers to apply these standards (a 'simplified CDISC' for COVID-19 may be useful).
4. In the current situation, standards related to data sharing around COVID-19 clinical research and case reporting should be made accessible without licensing fees. Openness should become the rule in pandemic situations.

FAIR Data

Discoverability and metadata are important elements to optimize sharing and accelerate data use. To prepare data for sharing clinical trial data should always be associated with adequate and standardized metadata to improve discoverability ("F" in FAIR).

1. 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 (e.g. ECRIN: Clinical Research Metadata Repository (CRMDR)[3].
2. Critical in the current situation is to have datasets easily findable. Resolvable persistent identifiers like DOIs, e.g. linking to a repository or network of repositories, would play a large part in making the data available.
3. For COVID-19 a variety of study designs is applied, covering interventional trials, observational studies, cohorts and registries. Metadata schemas between these study types should be aligned to improve discoverability of studies and associated data objects.

Protection of Trial Participants

1. Due to pressure to rapidly publish and make data available, there may be a greater risk of data not being properly de-identified (anonymized) prior to data sharing. For this reason, measures to protect and properly de-identify data is paramount (e.g. specific data use agreements). For public health emergency situations, some

legislations (e.g. GDPR Article 9 (2))[4] contain emergency provisions on processing of sensitive personal information in the area of public health, but even in this situation, the standard of protection of this data still requires safeguarding the rights and freedoms of the data subjects. This information should be available centrally on a government with explicit authority web page.

Informed Consent for Data Sharing

1. Data and clinical trial information should be made available for broad sharing when possible.
2. Individual participant data sharing should be based on broad consent by trial participants (or if applicable by their legal representatives) to the sharing and secondary reuse of their data for scientific purposes, according to applicable laws, regulations, and policies.
3. 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[4].
4. Procedures on data sharing specific for COVID-19 in the informed consent for clinical trials should be in accordance with standards and recommendations (e.g. ISO/TS 17975:2015 Health informatics - Principles and data requirements for consent in the Collection, Use or Disclosure of personal health information[1] or the Global Alliance for Genomics and Health (GA4GH) Consent Policy)

Publications and Other Formats

1. Availability for timely publication of results - even for negative and withdrawn studies - and for data underlying a publication 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). However, in the COVID-19 crisis, publication cannot be the criterion for data sharing. Timely data sharing should be performed as soon as the study is completed [10].
2. Pre-print publishing and other forms of knowledge sharing and exchange are also encouraged. Full reports should be made available immediately upon communication of results, e.g. through a press release.
3. Where possible, open access journals and adherence to OpenAire initiatives and the likes are also encouraged.

Credit and Attribution

1. In a situation where there is a strong need for data sharing, at an early stage and before primary papers might be written, having credit for the data becomes more important. Initiatives to support rewards and credits for data sharing should be strengthened (RDA: Sharing rewards and credit (SHARC) IG[5], FORCE11:Joint declaration of data citation principles[6])

2. Persistent Identifiers for data sources (e.g. DOI) should be included in a secondary analysis to recognize primary data providers.
3. Financial models to support data sharing for COVID-19 studies should be implemented and funding specifically targeted on such activities should be provided. This could include additional costs for preparing data sharing as well as making data as interoperable as possible.

Biological Samples as Data Sources

1. In the context of a pandemic the access to biological samples that are data sources might be of high interest and policies should be in place for facilitating their access; they should be developed in full respect of legal and safety regulations, protection of patients and with recognition of the value of the work performed to constitute such collections with relevant metadata and in line of the GDPR provisions on biobanking.[7]
2. Main principles are delineated in the Access policy of BBMRI-ERIC, the European research infrastructure consortium for biobanking and biomolecular resources[8].

Rights, Types and Management of Access

1. In order to expedite the process of data sharing, standardized agreements for sharing of data between data providers, repositories and data requestors for COVID-19 clinical trials should be developed and implemented (e.g. data transfer agreements, data access and data use agreements)
2. In the COVID-19 situation access to data should be as open as possible. This does not necessarily mean completely open access, as they also need to be as closed as necessary, but measures to control and manage risk (anonymization, aggregation, data use agreements) can be used to make access as easy as possible, while adequately protective. If a Data Access Board or a similar third-party mechanism is involved in decisions about data sharing, there is a need for a transparent and fast track process.

3.4 Guidelines

3.4.1 Clinical Trials on COVID-19

Clinical trials are an important research area to discover and make available safe and effective treatments for COVID-19[2] . International, regional, and national legal and methodological frameworks exist for clinical trials[3], 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 interventions are advisable.
4. Heterogeneity between 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. When regulatory bodies allow compassionate use of approved repurposed drugs such a use should be reported; if a fast track for approval of proved COVID-19 drugs exists it is also useful to report it. 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.

3.4.2 Other Types of Data

In COVID-19 clinical presentation and evolution, diagnostic and prognostic data including immunological data, virology tests results and imaging, especially lung scan in case of respiratory distress, are important elements. All values for metadata and assay results should be defined with the use of domain specific controlled vocabularies; a list of standards for clinical data and metadata is available at fairsharing. These data standards are recommended for the following data types:

1. Flow Cytometry (FACS) and Mass Cytometry (CyTOF) Experiments for ImmunoPhenotyping

The primary cytometry data in .fcs format is greatly enhanced by inclusion of interpreted data (e.g. the cell population name, definition and frequency). (https://www.immport.org/docs/standards/Cytometry_Data_Standard.pdf)

Cell population names should be the standard name from a curated reference source (e.g. Cell Ontology). Use of standardized Cell population names in flow cytometry and CyTOF experiments improves the ability to compare datasets.

Cell population definitions are based on the biomarker expression pattern or ‘gating strategy’. Biomarker names, when the biomarker is a monoclonal antibody, should use the antibody’s antigen name from Protein Ontology, UniProt, or ChEBI. Cell population frequency units should be defined. Inclusion of the monoclonal antibody's clone name enhances the confidence that this crucial assay reagent is the same across datasets.

2. Chemokine and Cytokine Measurements (e.g. ELISA, Luminex xMAP, MBAA)

Chemokine and cytokine assay methods are often based on monoclonal antibodies and findability and interoperability is facilitated by standardized naming of the antibody's antigen (e.g. Protein Ontology, UniProt, ChEBI), the antibody detector, the antibody's clone name and the vendor. Data standards and deposition guides are available (https://www.immport.org/resources/dataTemplates).

3. Neutralizing Antibody Titer

Standardized names for viral targets using reference sources (e.g. NCBI Taxonomy) is recommended. Description of the neutralizing antibody type (e.g. IgM, IgG) and detector enhances interoperability.

4. Virus Presence and Titer

Standardized names using reference sources (e.g. NCBI Taxonomy) for measurement of virus presence is recommended.

5. Imaging data

Standards for medical images and interoperability protocols such as those described in [11] should be applied. Digital Imaging and Communications in Medicine (DICOM) [12] — is *the* international standard for medical images and related information, that is universally adopted by almost all of the leading vendors of medical imaging equipment and software. Most relevant to COVID-19 is that virtually all clinical chest X-ray, lung CT, and brain/neuro MRI, and many ultrasound imaging systems follow the DICOM standard, which defines the formats for medical images that can be exchanged with the data and quality necessary for clinical use. A DICOM Tag serves as a unique identifier for an element of information which is used to identify Attributes and corresponding Data Elements. Supplement 142 of the DICOM Standard [13] offers a framework for de-identification of clinical imaging data for use in research studies. [14]

DICOMweb, a DICOM standard for web-based medical imaging, and HL7 FHIR are complementary standards to service the needs of imaging in healthcare. HL7 and FHIR provide the information model for health information, whereas DICOM and DICOMweb provide the information for imaging. [15]

4. Community Participation and Data Sharing

4.1 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.

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.

4.2 Scope

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.

This document is intended to look at data management and sharing issues and only reflect at the technical, social, legal and ethical considerations from that perspective.

Stakeholders

This document is intended to provide guidance and recommendations to the following groups of stakeholders.

1. Data subjects: Informed and forms of dynamic consent should be obtained from the data subject before personal data1 is collected from/about them and whenever there are changes to the data collection process, e.g. patients, citizens, general public.
2. Data processors/ data custodians/ data controller: determine the purposes and methods of the processing of personal data, perform the data processing, including analysis, anonymisation, storing and preservation, sharing e.g. researchers, app developer, funders, policy makers, health authority

What Do We Mean by App Development for Community-Generated Data? We are referring mostly to:

1. Symptom tracking apps (health monitoring apps where users self-report COVID-19 symptoms)
2. Contact tracing apps (mobile phone tracking used to identify potential geographic spread of COVID-19)
3. Services app (including service volunteers such as healthcare, shopping, entertainment, religious services)

Disclaimer: RDA does not endorse any products. Any products mentioned in this document is for illustrative purposes only, and does not constitute an endorsement by RDA. Please view the official RDA statement on this as referenced in the overall COVID-19 Guidelines and Recommendations section.

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.

4.3 Policy Recommendations

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.

4.3.1 Data Collection

App developers are not always aware of all the ethical and legal implications of the data they gather and might not be familiar with protocols for collecting and sharing data.

1. Ensure developers, data stewards, healthcare professionals, epidemiologists, researchers and the public are represented in the teams driving the development of the data collecting apps.

2. Consider the use of the data - clinical, social etc. This will help identify useful standards and disciplinary norms, provide additional directions on the necessary contextual information and harmonised metadata which will allow reuse and sharing across various information systems. Other sections of the RDA COVID-19 Guidelines and Recommendations provide guidance on some of these.

4.3.2 Data Quality and Documentation

In the race against time to collect the data required to combat the COVID-19 pandemic, there is risk that data is collected without sufficient attention to quality and reliability of data (e.g. level, or rather lack of any basic provenance of the data, quality of the sources, versioning and level of maintenance).Application developers may not always be aware of the required quality for data to be usable or reusable.

The research data community has been addressing these challenges, developing standards, vocabularies and ontologies, workflows and various disciplinary norms, as well as a key set of key principles to ensure data quality, findability, accessibility, interoperability and reuse (FAIR). Implementing the FAIR data principles will ease sharing and increase efficiency, especially important considering the time constraints we are facing.

4.3.3 Legal and Ethical Aspects

Ethical considerations have to be made regarding the two-way sharing of information using mobile-tracking apps.

1. Adequate medical, social and emotional support networks need to be established before apps relay to users they may have been in close proximity to a COVID-19 positive individual.The app project owners need to work with relevant local, national and international authorities to ensure appropriate support networks are in place and the app coordinates with these authorities in such matters. .

2. Make sensitive technical consideration such as transmitting anonymised codes as a means to alert individuals to exposure

4.3.4 Software Development

Contact tracing apps should adhere to the same development recommendations as other software, particularly to build public trust. While it has been highlighted that scientists must openly share the code behind modelling software so that the results can be replicated and evaluated (Barton et al. 2020), the transparency provided by open sharing can only address security concerns.

4.4 Guidelines

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.

4.4.1 Data Collection

1. Encourage public and patient involvement (PPI) throughout the data management lifecycle from inception of the research question, implementation of the data collection and final data sharing and usage.

2. Ensure apps are developed with the research and health care question as the central concept and only gather data needed to address these questions.

3. Applications designed to collect data should be developed as open source, with early release on a public code repository and made available under an open source licence (c.f. section on Research Software in this report), to build confidence in the public about security and privacy. It also allows for the rapid identification and removal of vulnerabilities.

4. Protecting personal data is of utmost importance when developing applications. Use protocols and methods that aim to protect personal data e.g. DP-3T.

4.4.2 Data Quality and Documentation

Follow standardised ways of collecting and curating community generated data and select a secure data collection platform and trusted digital repositories as a way of standardising COVID-19 data whilst ensuring quality and facilitating sharing. Compilation of recommended repositories can be found here and RDA recommends the use of CoreTrust approved repositories.

When collecting and curating the data, ensure detailed metadata is captured with the data. As a minimum the effort should be taken to include the following.

1. Protecting personal data is of utmost importance when developing applications. Use protocols and methods that aim to protect personal data e.g. DP-3T.

2. Provide contextual metadata to help processing, visualization, analysis, storage, publishing, archiving and reuse.

3. Include detailed descriptions of the methods, to aid verification of results.

4. Include details on the consent and type of consent associated with the collected data.

5. Metadata should also include any retention (and deletion) obligations associated with the data.

6. Also, where possible, consider including as metadata, specific information on technology characteristics and their limitations (eg: efficiency of underlying technology of app, eg: Bluetooth or GPS).

4.4.3 Data Storage and Long-Term Preservation

Considerations for long term storage and preservation of data generated from apps in relation to COVID-19 is not always apparent. For example, what are the retention periods that apply for COVID-19 related data? Due to the unprecedented nature of this pandemic, much of this is only being considered at present.

1. Ensure that all prevailing national and international legal and ethical requirements for health data and medical studies (e.g. for data retention periods) is adhered to.

2. Ensure that provision is made to enable easy updating of the data collection, storage and preservation to meet any changes to existing requirements.

3. Long-term preservation should be considered in the case of high-value data that could help in modelling future pandemics. Depositing the data in trusted and certified repositories that are widely used by the community aids in achieving this. FAIRsharing.org maintains a comprehensive catalogue of repositories which can be considered for this purpose.

4. Data should be available under an open licence that enables reuse, such CC-BY, unless there are legal and ethical considerations.

5. Consider benefits and challenges of either a centralised or decentralised model for data storage and processing. e.g. View the dedicated section on the processing in a centralised vs decentralised manner from the COVID SafePaths report "COVID-19 Contact-Tracing Mobile Apps: Evaluation And Assessment For Decision Makers" https://drive.google.com/file/d/1A9Ft7-YpB9IOCbaLrRHrR34XP2SiSet5/view

4.4.4 Transparency and Community Participation

Achieving a balance between timely contact tracing and community safety alongside individual privacy concerns such as surveillance, unauthorized use of personal data and forms of abuse that might result from the identification of subjects.

Establish appropriate and transparent governance mechanisms to have oversight of the data and its management. An open and transparent approach allows for the community to have a say and suggest improvements e.g. Guidelines from the Ada Lovelace Institute https://www.adalovelaceinstitute.org/our-work/covid-19/covid-19- exit-through-the-app-store/

5. Data Sharing in Epidemiology

5.1 Focus and Description

Responses to the COVID-19 pandemic have been massive and multifaceted worldwide. An immediate understanding of the disease’s epidemiology is crucial to slowing infections, minimizing deaths, and making informed decisions about when, and to what extent, to impose mitigation measures, and when and how to reopen society. Improved and innovative surveillance and follow-up across the globe is key to minimizing resurgence.

We are still in the midst of the current COVID-19 pandemic. Data and models are incomplete, provisional, and subject to correction under inconsistent and changing conditions. New data and newly applied analytics will provide a better understanding of our current situation, and new insights surrounding improved SARS-CoV-2 and COVID-19 responses. Despite the our reliance on, and the importance of evidence based policy and medical decisions, 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 quality 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, jpg), 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 due to divergent technical standards.

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 high quality data to support evidence-informed public health decisions during this rapidly evolving catastrophe.

5.2 Scope

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.

The RDA-COVID19-Epidemiology Work Group is working on how to improve practical responses, which means moving from conceptualization to realization, implementation, and execution (Figure 2). We have developed an Epidemiological Surveillance Data Model that identifies the primary data domains that need to be integrated to understand COVID-19, and to improve surveillance and follow-up. We propose the creation of a WHO-led COVID-19 *EPIdemiological Translational Research Action Coalition* (Epi-TRAC) to add an implementation layer to the existing WHO policies, guidelines, partnerships, and information exchange stack. A COVID-19 use case for Common Data Models (CDM) is introduced as a way to think about and start a discussion on implementation. Finally, we show how patient data can be re-used for research or public health purposes using anonymization and pseudonymization.

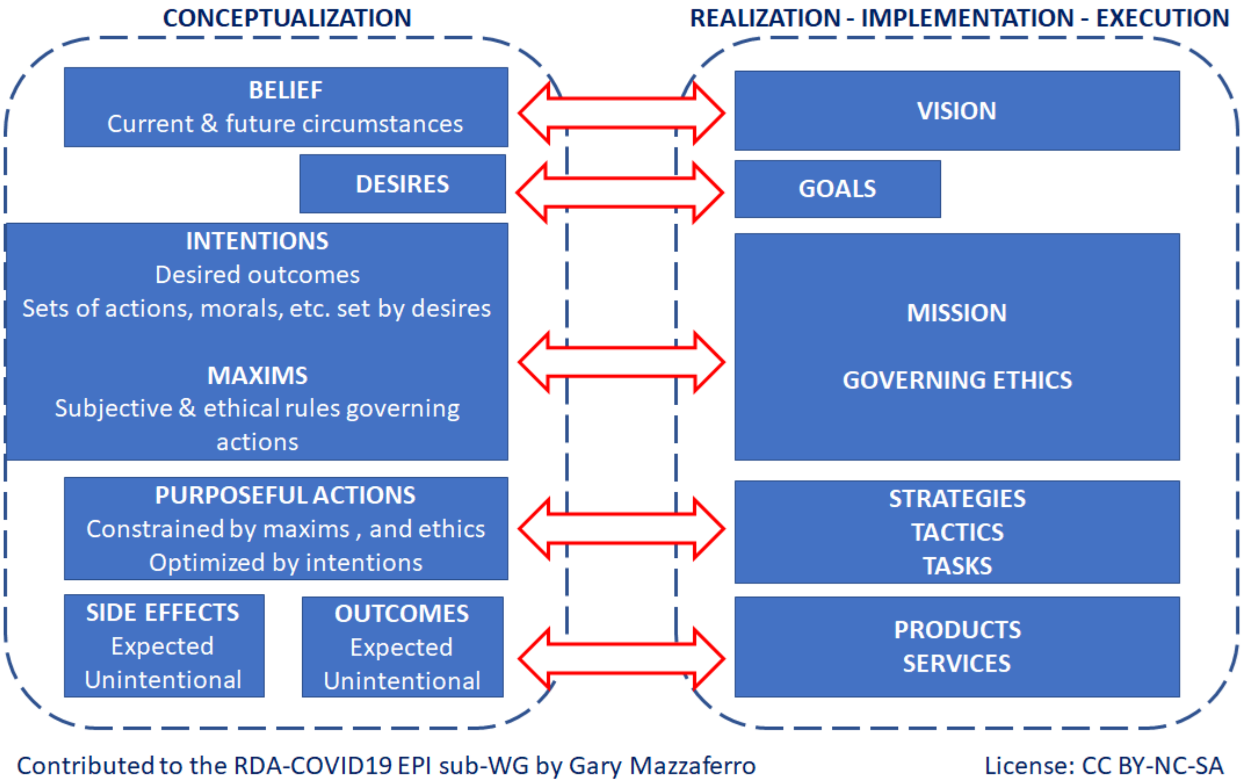

Figure 2. Moving from conceptualization to implementation (v0.01). Contributed by Gary Mazzaferro. LICENCE: CC BY-NC-SA 3.0

The intended audience for the epidemiology recommendations and guidelines are: government and international agencies, policy and decision makers, epidemiologists and public health experts, disaster preparedness and response experts, funders, data providers, teachers, and researchers.

5.3 Recommendations

5.3.1 Policy

1. Urgently update data sharing policies and Memoranda of Understanding (MOUs).
2. Implement a "data first" publication policy in research.
3. Accelerate the implementation of Open Data and Open Science tools and methods.
4. Add "Open Science" to the Open Government Partnership (OGP) list of policy areas.
5. Build and maintain public trust with policies of openness, transparency, privacy protection and honesty.

5.3.2 Information Technology And Data Management Infrastructure

1. Invest in information technology (IT) modernization, and interoperable data management system infrastructure.

5.3.3 Analysis and Modeling

1. Harmonized COVID-19 intervention protocols.
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.
4. Provide uncertainty quantification for all data and models.
5. Identify hotspots using data-driven approaches.

5.3.4 Surveillance Data

1. Develop consensus standard definitions and criteria for COVID-19 surveillance data across public health, clinical and other domains..
2. Document methodologies used to collect and compile data.
3. Develop standardised tools for aggregating microdata to harmonized formats.
4. Develop machine readable citations and micro-citations for dynamic data.

5.3.5 Global preparation, detection, and response

1. We propose the creation of a WHO-led COVID-19 EPIdemiological Translational Research Action Coalition (*Epi-TRAC*).

5.3.6 Interoperability and Data Exchange

1. Develop necessary technical specifications for record linkages across epidemiological input and output systems.

2. Develop systems to share pseudonymized data using encrypted person identifiers.
3. Share metadata and aggregated data where there are restrictions in accessing/using the related person-level data.

5.4 Guidelines

5.4.1 Policy

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*" publication policy in research 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).

5.4.2 Information Technology and Data Management Infrastructure

1. Invest 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.

5.4.3 Analysis and Modeling

1. Develop and implement internationally harmonized COVID-19 intervention protocols based on peer-reviewed empirical modeling and epidemiological evidence, taking into account local conditions.
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 utilisation, quality of life, social isolation. A key goal is to balance benefits due to reduced COVID-19 related mortality and morbidity with adverse side effects such as a wide range of bio-psychosocial and societal burdens.
4. Report underlying assumptions and quantify effects of uncertainties on all reported parameters and conclusions for all model predictions, data etc.
5. Implement a data-driven approach to identify hotspots.

5.4.4 Surveillance Data

1. Rapid development of a consensus standard on COVID-19 surveillance data:
    1.1. Definition of and reporting criteria for COVID-19 testing, reporting on testing, and testing turnaround times.
    1.2. Policies and definitions: interventions, contact tracing, reporting of cases, deaths, hospitalizations and length of stay, ICU admissions, recoveries, reinfections, time from contact if known, symptoms onset and detection, through clinical course and interventions, 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. Uniform standard 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 standardised tools for aggregating microdata to a harmonized format(s) that can be shared and used while minimising the re-identification risk for individual records.

4. Develop machine readable citations and micro-citations for dynamic data. 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 enough in the healthcare system nor via different survey instruments. Moreover, merging clinical databases with other types of databases is difficult or impossible due to interoperability and legal reasons.

Epidemiological Surveillance Data Model

Conceptualization of an epidemiological surveillance data model (**Figure 3**) 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. This is the domain model for COVID-19.

However, standardization challenges within each of these domains remain to be solved before data can be effectively integrated across domains for epidemiology studies. 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. Surveillance event history must be integrated across an interface 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 another 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.

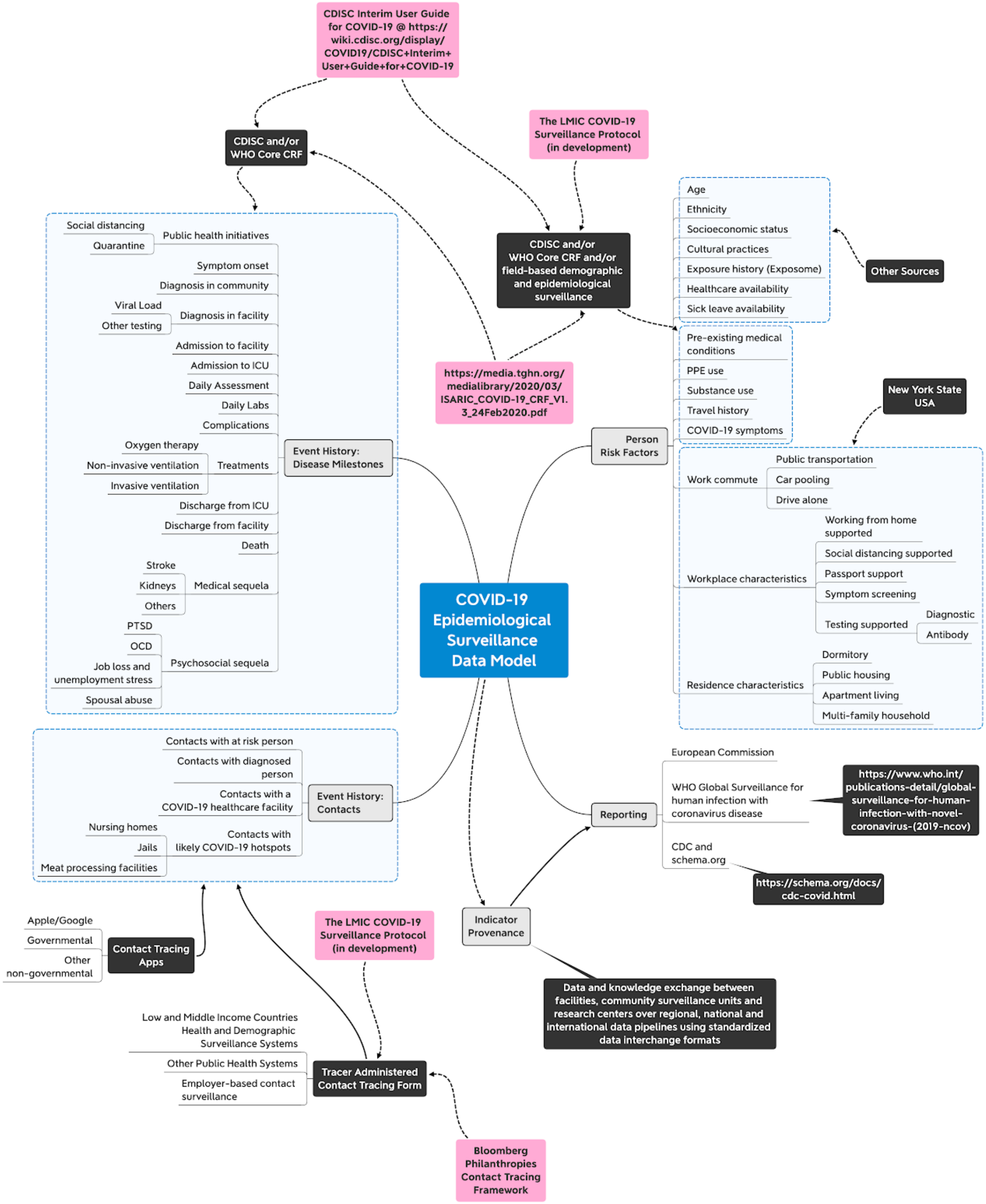
CDISC Interim User Guide for COVID-19 @ https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19
CDISC and/or WHO Core CRF
The LMIC COVID-19 Surveillance Protocol (in development)
CDISC and/or WHO Core CRF and/or field-based demographic and epidemiological surveillance
https://media.tghn.org/medialibrary/2020/03/ISARIC_COVID-19_CRF_V1.3_24Feb2020.pdf
Age
Ethnicity
Socioeconomic status
Cultural practices
Exposure history (Exposome)
Healthcare availability
Sick leave availability
Other Sources
Pre-existing medical conditions
PPE use
Substance use
Travel history
COVID-19 symptoms
New York State USA
Person Risk Factors
Work commute
Public transportation
Car pooling
Drive alone
Workplace characteristics
Working from home supported
Social distancing supported
Passport support
Symptom screening
Testing supported
Diagnostic
Antibody
Residence characteristics
Dormitory
Public housing
Apartment living
Multi-family household
Social distancing
Quarantine
Public health initiatives
Symptom onset
Diagnosis in community
Viral Load
Other testing
Diagnosis in facility
Admission to facility
Admission to ICU
Daily Assessment
Daily Labs
Complications
Oxygen therapy
Non-invasive ventilation
Invasive ventilation
Treatments
Discharge from ICU
Discharge from facility
Death
Stroke
Kidneys
Others
Medical sequela
PTSD
OCD
Job loss and unemployment stress
Spousal abuse
Psychosocial sequela
Event History: Disease Milestones
COVID-19 Epidemiological Surveillance Data Model
Contacts with at risk person
Contacts with diagnosed person
Contacts with a COVID-19 healthcare facility
Nursing homes
Jails
Meat processing facilities
Contacts with likely COVID-19 hotspots
Event History: Contacts
European Commission
WHO Global Surveillance for human infection with coronavirus disease
CDC and schema.org
Reporting
https://www.who.int/publications-detail/global-surveillance-for-human-infection-with-novel-coronavirus-(2019-ncov)
https://schema.org/docs/cdc-covid.html
Apple/Google
Governmental
Other non-governmental
Contact Tracing Apps
The LMIC COVID-19 Surveillance Protocol (in development)
Indicator Provenance
Data and knowledge exchange between facilities, community surveillance units and research centers over regional, national and international data pipelines using standardized data interchange formats
Low and Middle Income Countries Health and Demographic Surveillance Systems
Other Public Health Systems
Employer-based contact surveillance
Tracer Administered Contact Tracing Form
Bloomberg Philanthropies Contact Tracing Framework

Figure 3. Epidemiology surveillance data model (v0.02). Contributed by Dr. Jay Greenfield, developed from discussions during RDA-COVID19-Epidemiology Work Group meetings. LICENCE: CC BY-NC-SA 3.0. NOTE: We're working on fixing the resolution.

5.4.5 Global Preparation, Detection And Response

1. We propose the creation of a WHO-led COVID-19 EPIdemiological Translational Research Action Coalition (***Epi-TRAC***).

WHO’s Global Influenza Surveillance Response System (GISRS) is a well-established network of more than 150 national public health laboratories in 125 countries that monitors the epidemiology and virologic evolution of influenza disease and viruses (WHO 2020). On March 26, 2020, WHO published *Operational considerations for COVID-19 surveillance using GISRS*. This document,

> “*is intended for Ministry of Health and other government officials responsible for COVID-19 and influenza surveillance and summarizes the operational considerations for leveraging influenza surveillance systems to incorporate COVID-19 testing. The enhanced surveillance outputs will support national, regional, and global situation monitoring, knowledge building, risk assessment, and response actions*.”

Prior to the COVID-19 outbreak, WHO was already engaged in re-examining GISRS’s long-term fitness-for-purpose, reimagining the GISRS based on new themes. In line with these short-term considerations and with GISRS long-term aspirations, we are recommending a real time, adaptable, rapidly-responding system that supports developing countries, and that employs new technology to combat pandemics and other emerging diseases. The RDA-COVID19-Epidemiology WG recommends the creation of a WHO-led *Epi-TRAC* to add an implementation layer to the existing WHO policies, guidelines, partnerships and information exchange stack (Figure 4).

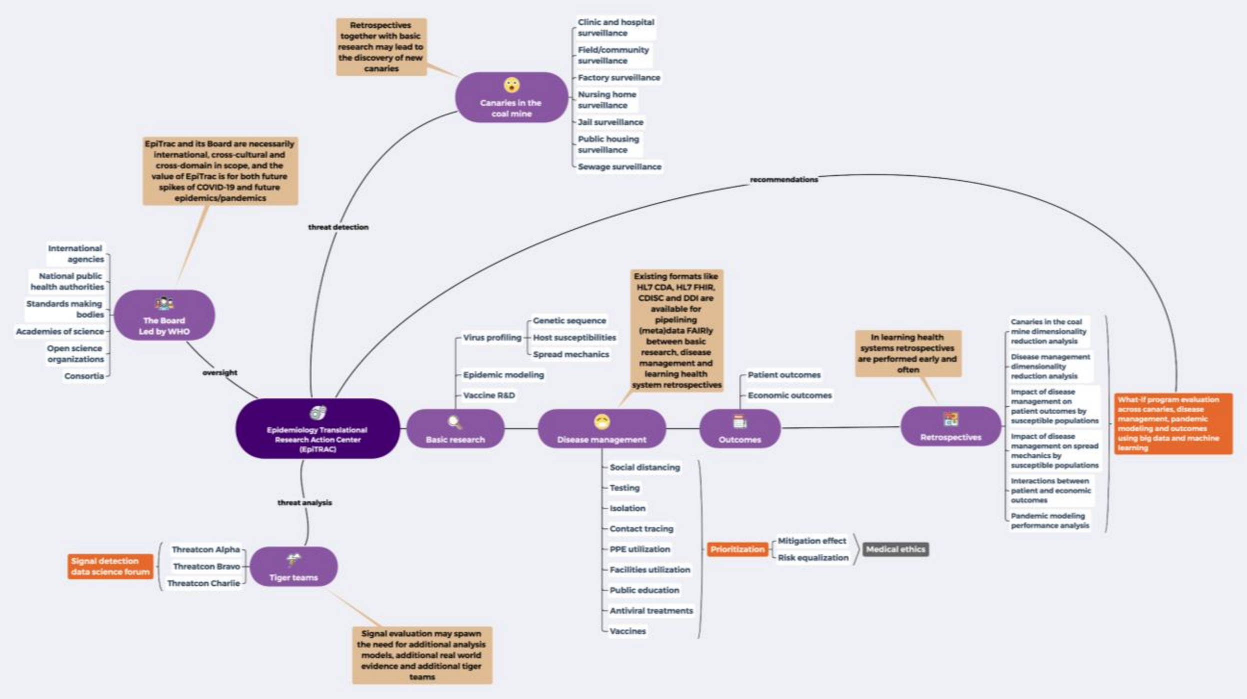
Retrospectives together with basic research may lead to the discovery of new canaries
Canaries in the coal mine
Clinic and hospital surveillance
Field/community surveillance
Factory surveillance
Nursing home surveillance
Jail surveillance
Public housing surveillance
Sewage surveillance
EpiTrac and its Board are necessarily international, cross-cultural and cross-domain in scope, and the value of EpiTrac is for both future spikes of COVID-19 and future epidemics/pandemics
threat detection
recommendations
International agencies
National public health authorities
Standards making bodies
Academies of science
Open science organizations
Consortia
The Board Led by WHO
oversight
Epidemiology Translational Research Action Center (EpiTRAC)
Basic research
Virus profiling
Genetic sequence
Host susceptibilities
Spread mechanics
Epidemic modeling
Vaccine R&D
Existing formats like HL7 CDA, HL7 FHIR, CDISC and DDI are available for pipelining (meta)data FAIRly between basic research, disease management and learning health system retrospectives
Disease management
Social distancing
Testing
Isolation
Contact tracing
PPE utilization
Facilities utilization
Public education
Antiviral treatments
Vaccines
Prioritization
Mitigation effect
Risk equalization
Medical ethics
Outcomes
Patient outcomes
Economic outcomes
In learning health systems retrospectives are performed early and often
Retrospectives
Canaries in the coal mine dimensionality reduction analysis
Disease management dimensionality reduction analysis
Impact of disease management on patient outcomes by susceptible populations
Impact of disease management on spread mechanics by susceptible populations
Interactions between patient and economic outcomes
Pandemic modeling performance analysis
What-if program evaluation across canaries, disease management, pandemic modeling and outcomes using big data and machine learning
threat analysis
Signal detection data science forum
Threatcon Alpha
Threatcon Bravo
Threatcon Charlie
Tiger teams
Signal evaluation may spawn the need for additional analysis models, additional real world evidence and additional tiger teams

Figure 4. Epi-TRAC (v0.02). A proposed WHO-led COVID-19 EPIdemiological Translational Research Action Centre. Contributed by Dr. Jay Greenfield and Gary Mazzaferro, developed from discussions held during RDA-COVID19-Epidemiology Work Group meetings. Special thanks, also to Dr. Stefan Sauermann, and Gary Mazzaferro. LICENCE: CC BY-NC-SA 3.0

5.4.6 Interoperability and Data Exchange

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 complete Metadata in machine readable formats where there are restrictions in accessing/using the related data.

Anonymization and pseudonymization are tools that can be used to enable cross-border data discovery and data transfer while at the same time ensuring compliance with privacy requirements. This is essential during the COVID-19 pandemic when data sharing is essential to enable epidemiological analysis, cross-border pandemic modeling, and coordinated policy development between countries.

5.4.6.1 COVID-19 Questionnaire Initiatives

A use case for anonymization, pseudonymization, cross-border data discovery and cross-border data transfer are the international efforts underway to create interoperable COVID-19 demographic and epidemiological surveillance questionnaires. These initiatives are domain-specific and support the localization of questions and answers in order to assure cross-country and cross-cultural comparability.

There are a number of actively developing COVID-19 questionnaire initiatives that figure into one or more international efforts to create interoperable COVID-19 epidemiological surveillance questionnaires (**Tables 1 and 2**). These initiatives are very much a work in progress at this time. Indeed, some of us are already participating in these international efforts and are just now learning about each other's work via the RDA-COVID19-Epidemiology Work Group.

Table 1 Questionnaire instruments: Reference studies. Development of such initiatives is a very active and rapidly changing area at the present time during the COVID-9 pandemic. Contributed by Dr. Carsten Schmidt, Dr Jay Greenfield, Dr. Stefen Sauermann, and Dr. Chifundo Kanjala, developed from discussions held during RDA-COVID19-Epidemiology Work Group meetings

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
1 Australia	NSW Case questionnaire	Patients		English		
2 Brazil	Brazil Prevalence of Infection Survey	Rapid tested, Tested positive	In development	English		
3 Europe	Questionnaire by WHO Europe	General population		German, Russian		Single Instrument

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
France	Barometer Covid19					The "barometer Covid19" is a citizen science initiative driven by the Datacovid association. It aims to provide open-access data from a weekly survey to illuminate the struggle against the epidemic Covid19 from observations on its dynamics, its determinants and its impacts.
France	French COVID-19	Patients	In use	English		Lead by REACTing consortium in collaboration with ISARIC consortium (International Severe Acute Respiratory and emerging Infection Consortium)
4 Germany	Covid-19 research dataset	Patients	In development	German		National Network of German University Clinics to study COVID19.

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
6 Germany	GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany	General population	In use	German		As the largest European infrastructure institute for the social sciences GESIS provides essential and internationally relevant research-based services
7 Israel	One-minute population wide survey (Israel)	Isreali population	In use	Hebrew, Arabic, Russian, Spanish, French, English		Participants asked to fill it out on a daily basis and separately for each family member, including members who are unable to fill it out independently (e.g., children and older people).
8 Low and Middle Income Countries	LMIC Covid Questionnaire		In development		UK COVID-19 questionnaire, SAPRIN COVID-19 screening form	
9 South Africa	South African Population Research Infrastructure (SAPRIN) COVID-19 Screening Form		In development	English, Afrikaans		

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
10 Uganda	Perinatal COVID-19 Uganda	Women pre-/perinatal		English		
11 UK	UK COVID-19 Questionnaire	Adult Respondent, Children, Key worker, Partner.	In development	English	NIHR Global Health Research Unit	- World Bank code book for metadata. - Becoming a model for some African countries
12 UK	National Institute for Health Research (NIHR) Global Health Research Unit	Telephone sample		English		
13 US	Human Infection with 2019 Novel CoronavirusPerson Under Investigation (PUI) and Case Report Form	Patients	In use	English	CDC	

Country	Initiative	Target population	Development stage	Language	Provenance (Influenced by...)	Comments
14 US/WHO	Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection	Patients	In use	English	WHO	This protocol has been designed to investigate the extent of infection, as determined by seropositivity in the general population, in any country in which COVID-19 virus infection has been reported.

Footnotes (Domains)

1. Clinical symptoms, Disease outcome, Exposure sites, Pre-existing conditions, Risk history, Sociodemographics.
2. Demographics, Home life, Test results, Transportation
3. Affect, Behaviour, Conspiracies (perceptions), COVID-19 risk perception: probability and severity, Fairness (perceptions), Frequency of Information, Influenza risk perception: probability and severity,, interventions (perceptions), Knowledge and self-assessed adherence to prevention measures, Knowledge incubation, Knowledge symptoms/treatment, Lifting restrictions (pandemic transition phase), Policies, Preparedness and perceived self-efficacy, Prevention – own behaviours, Resilience (perceptions), Risk group, Rumors (open-ended), Self-assessed knowledge, Socio-demography, Trust in institutions (perceptions), Trust in sources of information, Use of sources of information, Worry.
4. Clinical symptoms, Complications, Imaging, Laboratory markers, Medical treatments, Medication, Sociodemographics.
5. Adherence to risk minimization measures, Changes in lifestyle factors, COVID infections and testing, COVID tracing, General health status, Physical symptoms, Respiratory infections, Workplace/changed employment situation.
6. Changed employment situation, Childcare obligations, Risk perception, Evaluation of political measures & their compliance, Media consumption, Risk minimization measures, Trust in politics and institutions.
7. Age, Geographic location (city and street), Isolation status, Sex, Smoking habits
9. Actions in response to COVID-19, Bounded structure, Eligibility for testing, Epidemiological risk, Household enumeration, Household impact, Quarantine and hygiene directions,, Symptom screen, Travel and movement (mobility), Travel history, Visit attempts.
10. Disease outcome, Sociodemographics, Symptoms
11. Accommodation type, Away from home environment, Behavior changes, Change in benefits, Digital access, Economic activity before and after lockdown, Environmental attitudes, Environmental impact, Family relations, Financial impact, Food security, Impact on employment, Knowledge, Medication, Mental health, Mental health, Physical health, Pre-existing conditions, Social impact, Symptoms, Volunteering.
12. COVID-19 Interventions, Current living conditions, Displacement and mobility, Economic impacts, Impact of COVID-19 on health-related behaviors, Mental health, Precautions, Pre-existing conditions, Social aid, Social impact, Symptoms, Treatments for pre-existing conditions.
13. Diagnostic testing procedures, Clinical course, Medical history, Pre-existing conditions, Risk exposure, Sociodemographics, Symptoms, Treatments.
14. Laboratory results, Sociodemographics, Symptoms.

Table 2 Questionnaire instruments - Resources. Development of such initiatives is a very active and rapidly changing area at the present time during the COVID-9 pandemic.

Provider	Initiative	Language
US	NIH Public Health Emergency and Disaster Research Response (DR2)	Diverse
NIH	COVID-19OBSSR Research Tools	Diverse
PhenX is funded by the National Institutes of Health (NIH) Genomic Resource Grant	PhenX COVID-19 Toolkit	Diverse

Some of the questionnaire initiatives shown in Tables 1 and 2 are currently feeding into the construction of a COVID-19 demographic and epidemiological surveillance question bank.

Note that at the time of writing of the present document, Tables 1 and 2 are targeted for improvements in several respects:

1. Instruments and other resources that have been identified require additional review to ensure that key initiatives for COVID-19 have not been overlooked;
2. Consistent categorization of domains and cohorts will increase the usefulness of Table 1 and 2;
3. More provenance information could prove useful to researchers seeking to understand the effects of the pandemic internationally; and,
4. Additional comments and/or contextual information could provide tips about what these initiatives do well and what improvements they might make.
5. How to ensure adequate data to answer research questions rigorously, unambiguously and transparently.

5.4.6.2 COVID-19 Question Bank

The Wellcome Trust is participating in the development of a COVID-19 question bank that can be used to form locality specific surveys with both common and distinct questions by domains and cohorts.

Specific initiatives such as the UK COVID-19 Questionnaire, and the Low and Middle Income (LMIC) Questionnaire for Sub-Saharan Africa and Asia (under development) are now being funded. Such funding may kick start the development of a domain- and cohort-specific question bank. This question bank, once it becomes operational, can, in turn, be queried and filtered by domain, cohort, question text and so forth. Based on such queries, new questionnaire products can be developed that are more or less interoperable, depending on the questions selected and the capture of "localization" information in the question metadata when questions are reused from one survey to the next.

Reporting formats vary considerably between governmental agencies, non-governmental agencies, and the various communities of practice. A translation ecosystem needs to grow around reporting formats to facilitate frictionless flow of information during a pandemic. This ecosystem is indicated by the "exchange" arrow at the "disseminate" box in Figure 5.

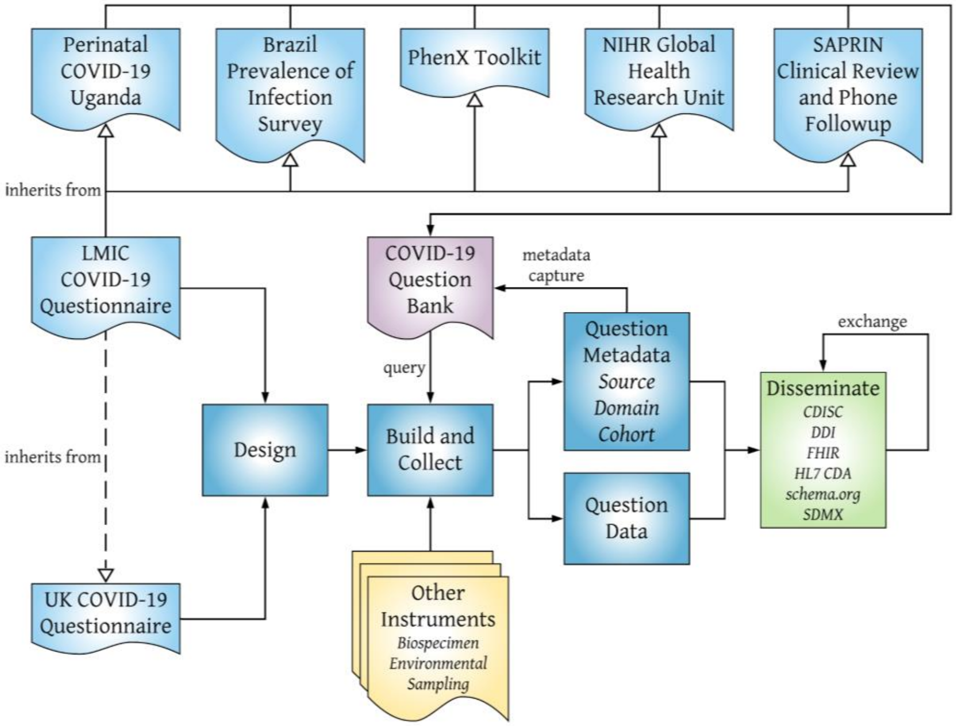

Figure 5. A draft model domain- and cohort-specific COVID-19 question bank to facilitate frictionless flow of information during the pandemic (v0.02). See, also, Tables 1 and 2 Contributed by Dr. Jay Greenfield, Dr. Chifundo Kanjala, and Dr. Carsten Schmidt, developed from discussions held during RDA-COVID19-Epidemiology Work Group meetings and based on work in the U.K., Asia, and sub-Saharan Africa supported by the Wellcome Trust. LICENCE: CC BY-NC-SA 3.0

Cross-border, cross domain and semantically interoperable data sources are key to sharing and linking data for pandemic policy making. 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 patient care. 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 optimise 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. optimising 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, CDISC must be considered as well as other technical standards from the more clinical space (IHE and HL7). In parallel to short term activities, long-term cooperation needs to be established, under clear coordination and with sufficient resources.

5.4.6.3 COVID-19 Use Case for Common Data Models

Epidemiological surveillance data collected in the field using questionnaires is just one player in an ecosystem of COVID-19 data that can include contact tracing apps, biospecimen and environmental sample data collected in the field and patient care data collected in clinics and hospitals.

Consider three geographic regions or cohorts. One captures patient care data manually using the WHO COVID-19 Rapid Version CRF. The second cohort uses one EHR system based on the HL7 CDA standard. A third cohort uses an EHR system based on the HL7 FHIR standard. Both of the cohorts using EHR systems produce a COVID-19 Rapid Version CRF as a report.

Consider next that each of these three regions wish to combine patient care data with epidemiological surveillance data in order to gain a more complete picture of the state of the pandemic in their region in line with the latest and greatest COVID-19 epidemiological surveillance model. Each region plans to use this more complete picture to create an early warning and response system for public health policy makers.

However, just like with patient care, each region has taken a set of questionnaire specifications and built the questionnaires using questionnaire systems that use different standards and formats for representing and storing the answers.

Based on this use case, cross-border interoperability of COVID-19 data is a hard problem. In fact, it was a hard problem before COVID-19. All COVID-19 has done is to reveal anything and everything that is weak or broken when we really need it. That being said here are some ways we propose to fix this problem slowly, based on previous initiatives, some ingenuity and a lot of expediency.

There are old experiments and new ones under way that attempt to host patient care and epidemiological surveillance data together in a Common Data Model. For example, "the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) has been shown to be an effective way to standardize observational health databases but has not been as commonly applied to survey and registry databases as it has for electronic health records and administrative claims”. In an experiment the NHANES dataset was successfully transformed into the OMOP CDM “using certain methods”. Note that, “the National Health and Nutrition Examination Survey (NHANES) is a program that combines survey information and physical examination results to determine the prevalence of major diseases and risk factors for disease among the U.S. population”. In this respect, NHANES is not unlike the HIC and LMIC COVID-19 questionnaires referenced above. See this poster which summarizes articles from two journal publications for a description of the experiment and details about the “certain methods” it uses. They are the secret sauce.

Generally speaking, before questionnaire content can be useful alongside patient care information, its subject matter needs to be mapped to domains and concepts. Domains make the subject matter findable and concepts play a hand in making questions interoperable.

In questionnaires it is necessary to know the concept behind a question for many reasons:

1. Questions and their response categories can vary from one country to the next and we need to assess whether they are measuring the same thing
2. Questions and their categories may change over time and we need to assess whether there is “drift” or, alternatively, they are measuring the same thing
3. In the context of patient care and when we collect biospecimens in the field for analysis in a lab and/or storage in a biobank. we need to know whether in patient care and in epidemiological surveillance, upon the results, the same concepts are being measured

EHR systems automatically associate concepts with the observations they make. However, when multiple EHR systems are in play across borders or with certain standalone CRFs, observations may not come with concepts or the mapping between questions and concepts may become clouded.

The VODAN initiative fills this gap with the WHO COVID-19 Rapid Version CRF. It provides a semantic data model for the COVID-19 Rapid Version CRF that associates CRF questions/items with domains and concepts.

In pushing questionnaire data into Common Data Models alongside patient care data, humans assisted by Natural Language Processing (NLP) or NLP assisted by humans have proven to be valuable when it comes to mapping questions to concepts.

COVID-19 is a good use case for Common Data Models because we want to contrast and compare patient care approaches for cohorts with different environmental exposures over time. The concept behind “different environmental exposures” is the “exposome”. Note that the exposome “comprises the totality of exposures to which an individual is subjected from conception to death, including those resulting from environmental agents, socioeconomic conditions, lifestyle, diet, and endogenous processes” (from the Dictionary of Epidemiology MS Porta, 6th edition, OUP 2014)

The COVID-19 use case for Common Data Models enables us to assess the outcomes of different patient care approaches for cohorts with different exposomes in emulated clinical trials. For a discussion of comparative effectiveness research, common data models and emulated clinical trials, see "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available", Hernan MA and Robins JM, Am J Epidemiol. 2016 Apr 15;183(8):758-64. doi: 10.1093/aje/kwv254. Epub 2016 Mar 18.

5.4.6.4 Anonymization and Pseudonymization

Patient related data are recorded and used in different domains, for example medicine, communities, research, administration, and statistics. Patient specific electronic identifiers (IDs) link specific information to the individual patient records in each domain. Each domain assigns a domain specific ID to each individual patient. In order to satisfy privacy requirements, data that carry a domain ID remain within the home domain, and are not shared.

If data are re-used e.g. for research or public health purposes, the domain specific ID is removed (anonymization) or replaced by a different ID (pseudonym). Pseudonyms can later be traced back to the original domain ID. This must only occur under well-defined conditions, e.g. if a statistics department needs to clear ambiguities in incoming information together with the organization that generated the data. In multi-domain and multi-organization scenarios, consistent management of IDs and pseudonyms is needed to enable cross-linking of data from different sources while ensuring privacy.

The following requirements apply:

1. Patient IDs exist for each domain;
2. The local domain ID must not leave the local domain in clear text, to prevent unintended record linkage between domains;
3. When providing data from a source domain to a target domain, the target domain patient ID (pseudonym) must become available to users in the target domain; and,
4. Domain IDs must enable domains to cooperate e.g. for clearing ambiguities, while preserving privacy and pseudonymization

In Austria, for example, eGovernment legislation and IT infrastructures are in place to handle domain specific identifiers e.g. for health care, traffic, taxes, and statistics (reference). This is implemented and in operation for example in the Austrian electronic health care record ELGA (reference). Figure 6 describes how data from a health domain can be linked to records in a research domain in this way. Figure 7 introduces the needed IT infrastructure. Figure 8 shows how IDs are mapped between domains while preserving pseudonymization.

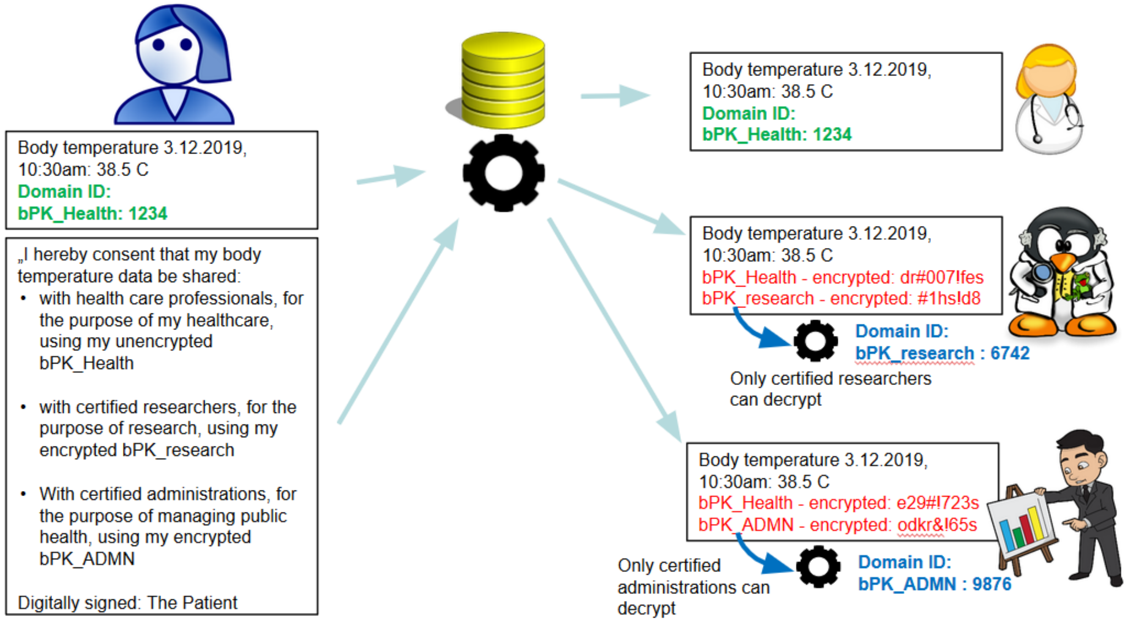
Body temperature 3.12.2019,
10:30am: 38.5 C
Domain ID:
bPK_Health: 1234
„I hereby consent that my body temperature data be shared:
• with health care professionals, for the purpose of my healthcare, using my unencrypted bPK_Health
• with certified researchers, for the purpose of research, using my encrypted bPK_research
• With certified administrations, for the purpose of managing public health, using my encrypted bPK_ADMN
Digitally signed: The Patient
Body temperature 3.12.2019,
10:30am: 38.5 C
Domain ID:
bPK_Health: 1234
Body temperature 3.12.2019,
10:30am: 38.5 C
bPK_Health - encrypted: dr#007!fes
bPK_research - encrypted: #1hs!d8
Domain ID:
bPK_research : 6742
Only certified researchers can decrypt
Body temperature 3.12.2019,
10:30am: 38.5 C
bPK_Health - encrypted: e29#!723s
bPK_ADMN - encrypted: odkr&!65s
Only certified administrations can decrypt
Domain ID:
bPK_ADMN : 9876

Figure 6. Sharing or linking a body temperature observation from the healthcare domain with a research and administration domain. In the healthcare domain ID (bPK_Health, green), a patient is identified with a specific ID, (1234, colour green, denoting that it is unencrypted). A doctor will receive this data together with the original ID, as the law allows doctors to share IDs unencrypted. The doctor can attach an encrypted ID (#1hsId8, red, denoting it is encrypted) to the data. A researcher who receives the data, decrypts the encrypted ID. This decrypted ID (6742, blue, denoting that it is a pseudonym) is specific to the research domain (bPK_research, blue). The same method applies as data are provided to administrations, e.g. for public health purposes. Contributed by Dr. Carsten Schmidt. LICENCE: CC BY-NC-SA 3.0

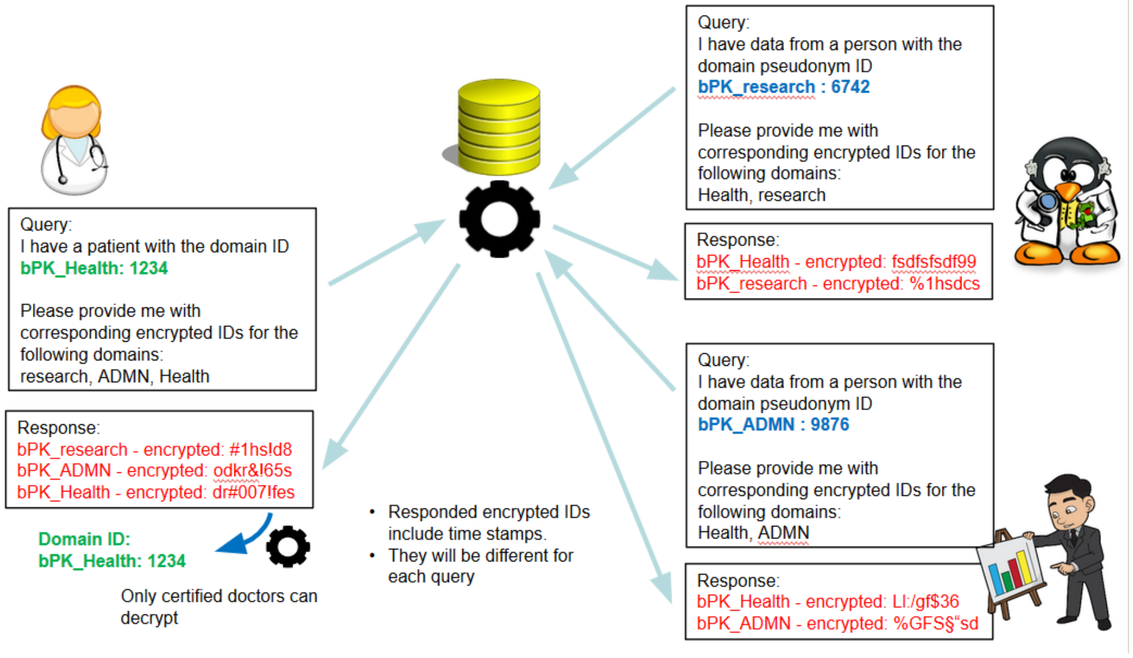
Query:
I have a patient with the domain ID
bPK_Health: 1234
Please provide me with corresponding encrypted IDs for the following domains:
research, ADMN, Health
Response:
bPK_research - encrypted: #1hs!d8
bPK_ADMN - encrypted: odkr&!65s
bPK_Health - encrypted: dr#007!fes
Domain ID:
bPK_Health: 1234
Only certified doctors can decrypt
• Responded encrypted IDs include time stamps.
• They will be different for each query
Query:
I have data from a person with the domain pseudonym ID
bPK_research : 6742
Please provide me with corresponding encrypted IDs for the following domains:
Health, research
Response:
bPK_Health - encrypted: fsdfsfsdf99
bPK_research - encrypted: %1hsdcs
Query:
I have data from a person with the domain pseudonym ID
bPK_ADMN : 9876
Please provide me with corresponding encrypted IDs for the following domains:
Health, ADMN
Response:
bPK_Health - encrypted: Ll:/gf$36
bPK_ADMN - encrypted: %GFS§“sd

Figure 7. IT infrastructure for cross-domain IDs management. A user in the medical domain queries the IT service using the ID of the health domain, asking for encrypted IDs of other domains. The service responds with the encrypted IDs. Users in other domains can use the same mechanism. This enables users in different domains to co-

operate: For example the researcher can attach the encrypted bPK_Health ID to a message to the doctor, asking for details to support clearing ambiguities in the data the doctor provided earlier. The doctor can then decrypt the patient ID, access the patient related information in the health IT system, and finalise the clearing with the researcher. Contributed by Dr. Carsten Schmidt. LICENCE: CC BY-NC-SA 3.0

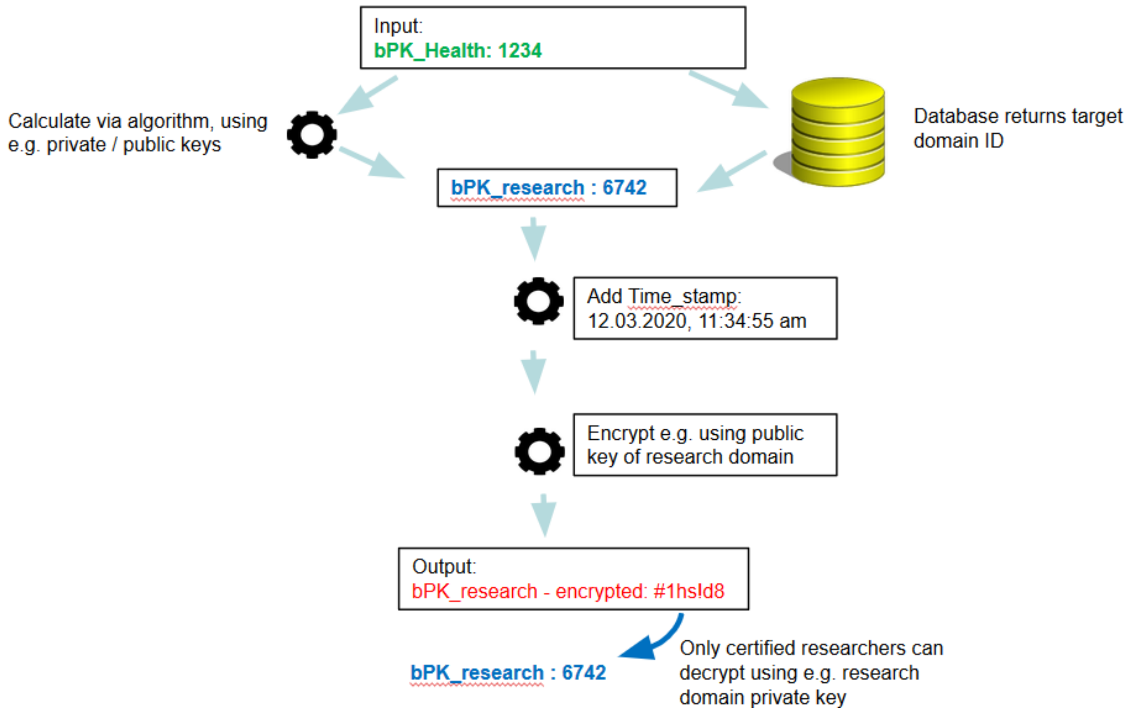
Input:
bPK_Health: 1234
Calculate via algorithm, using e.g. private / public keys
Database returns target domain ID
bPK_research : 6742
Add Time_stamp: 12.03.2020, 11:34:55 am
Encrypt e.g. using public key of research domain
Output:
bPK_research - encrypted: #1hs!d8
bPK_research : 6742
Only certified researchers can decrypt using e.g. research domain private key

Figure 8. Data flow for deriving an encrypted ID (#1hs!d8) for a target domain (research). The source ID (1234) can be mapped to the target identifier for example in two ways. A mathematical algorithm is used (left branch) to calculate the target domain ID (6742), or a database query returns the ID. A time stamp is then attached to this ID. ID and timestamp together are then encrypted, e.g. using the public key of the target domain. This assures that no two encrypted IDs for the same patient and the same domain are identical, in this way preventing the unintended linking of records. Contributed by Dr. Carsten Schmidt. LICENCE: CC BY-NC-SA 3.0

6. Data Sharing in Omics Practices

6.1 Focus and Description

The understanding of the ways in which the SARS-CoV-2 virus causes the COVID-19 disease is based on research into the molecular biology of the processes at cellular and subcellular level. The data of this style are the focus of this section.

6.2 Scope

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 biological research on SARS-CoV-2 and COVID-19.

6.3 Policy Recommendations

6.3.1 Researchers Producing Data

1. The FAIR data principles (Wilkinson et al. 2016) address a primary concern that has led to the formation of the group writing these guidelines: availability and re-usability of research data on COVID-19, in order to prevent unnecessary duplication of work. Many of the specific guidelines in this chapter (and others) address what can be done to make the data as FAIR as possible with a reasonable time investment. Some considerations during the COVID-19 pandemic are:

   1.1. Several of the FAIR principles call for rich metadata. Especially where data about human subjects is concerned it is not always possible to share such metadata in an open catalogue. Specifics can be found in guidelines for the individual data types as well as in the chapter on legal issues and ethics.

   1.2. In this time of urgency, making data FAIR should not unnecessarily slow down researchers collecting data. It is better to bring researchers collecting data with data stewards who can help with the FAIRification than to force everybody to learn how to do this themselves.

   1.3. We also need to encourage people to share what they have as-is without fear it is insufficient, and signal that help is needed

   1.4. A generic guideline for increasing FAIRness is to make sure data is made available in existing (certified, e.g. (CoreTrustSeal Standards and Certification Board 2019)) repositories, rather than starting new local resources. Also, if a choice must be made, submission to domain-specific repositories is preferred over generic repositories and catalogs.

   1.5. Reusability of data requires documented provenance: When sharing any secondary data the generation of which involved comparison against other resources (examples for OMICS data are: reference sequences for mapping, GO annotations for expression analysis, pre-trained models for gene annotation), both the public availability of these used resources and unambiguous referencing of the used resources, including version numbers, should be ensured.

1.6. Increase the reusability of data with consistent preprocessing: To increase the availability of data ready for analysis and integration, it may be prudent to agree on a consistent approach to preprocessing OMICS data. This would be a second-phase step that should not unnecessarily slow down researchers collecting data.

2. The FAIR principles do not contain a push for open availability of the data, but they are often accompanied by the credo "as open as possible, as closed as necessary". In these times of the pandemic, this quest for openness gains even more importance. It is therefore critical to pursue "legal interoperability", which in this context practically means to use a CC0 waiver where possible, a CC BY 4.0 license if necessary, with a strong preference for not adding any other restrictions (RDA-CODATA Legal Interoperability Interest Group 2016). For more details on this, we refer to the guidelines of the Legal/Ethics WG.

3. Data reproducibility and increased trust in the shared data are important. This is covered in detail in the section 8 "Research Software and Data Sharing" Summarizing:

    3.1. Software, including scripts and applications that were used to process and analyse the data should be provided along with the data in the publication.

    3.2. The dependencies of the underlying software environment should also be provided with the data in the publication.

A good example of these principles in action is the Galaxy platform (Galaxy Project 2020).

6.3.2 Funders

1. It is recommended that increased weighting is given in the grant review process to researchers who demonstrate best practice in open data and data reproducibility with their research outputs.

2. Require association of a project with a dedicated data steward, who can be specifically responsible for making data available in a FAIR and timely manner without interfering with the required pace of the research process.

3. Make sure that calls for projects clearly say that for COVID-19 data "timely" publication means "as soon as possible after it has been collected" and not "as soon as the publication has been accepted by the journal".

4. Be clear in the call for proposals that budget for professional data stewards in the project to help make the data more FAIR is eligible for funding.

5. Due to the high costs involved with high-throughput genomics, little data is available from Low and Middle Income Countries (LMICs) and from minority ethic populations in high income countries, thus leading to improper extrapolation of results to unrepresented population groups. Research that improves the coverage could be worth preferential treatment for funding.

6.3.3 Publishers

1. Require publishing of data underlying a study, in an even more timely manner than usual.

2. Make sure that the author recommendations prefer publishing of data in domain-specific repositories where findability is better than in generic or institutional repositories.

6.3.4 Policymakers

1. Put guidelines into place that give researchers ease of mind when licensing/sharing their data.

2. Promote use of domain-specific repositories instead of or as well as institutional repositories. Benefits of domain-specific repositories are that metadata standards are more likely enforced, assay result formats that promote re-use better supported, standardization of terms and ontologies eases re-use challenges.

3. Due to the high costs involved with advanced high-throughput genomics, data is disproportionately not available from Low and Middle Income Countries (LMICs) and from minority ethic populations in high income countries, thus leading to improper extrapolation of results to unrepresented population groups. A strong policy framework is required to facilitate research and encourage more inclusive participation.

6.3.5 Researchers

1. If you have any existing SARS-CoV, MERS-CoV or EBOV data that have not yet been made public, consider publishing that data now as it can be a useful reference.

2. Think early about systematic naming of filenames. Not thinking about it early enough is often the cause of a lot of extra work when the data is not stored in a database and researchers have to rename a large number of files manually at a later stage.

3. Document the computing time and resources required for data processing. This could help other researchers to assess the time and resources required for the pipeline, therefore to decide whether it is feasible to proceed with the local resources available.

4. When selecting a repository for submission of the data, priority should be given to domain-specific repositories over generic (e.g. institutional) repositories. Domain-specific repositories are easier to find, and often have better visualization and selection facilities for re-users of the data.

5. The repositories listed for deposition are also prime locations for locating existing data. Many now have dedicated sections for new as well as pre-existing data relevant to COVID-19 research.

6.3.6 Providers of Data Sharing Infrastructures

Perform validation that data confirms to recommended metadata/annotation standards in order to help researchers making their data as FAIR as possible.

6.4 Guidelines

This chapter addresses guidelines that are appropriate for all OMICS data types, and potentially also for data addressed in other chapters.

6.4.1 Recommendations for Virus Genomics Data

6.4.1.1 Repositories

There are several genomics resources that can be used to make virus genomics sequences available for further research. A curated list can be found in FAIRsharing (in FAIRsharing). Some specific examples are:

1. We suggest that raw virus sequence data is stored in one of the INSDC archives (INSDC, n.d. http://www.insdc.org/), as each of these is well known and openly accessible for immediate reuse without undue delays:

    1.1. DDBJ (Ogasawara O, 2020; in FAIRsharing) Sequence Read Archive

    1.2. ENA (European Nucleotide Archive at EMBL-EBI; in FAIRsharing), for submission documentation see ENA Documentation (ENA-Docs, 2020)

    1.3. NCBI SRA (in FAIRsharing), for submission documentation see SRA Submission documentation (NIH-NCBI, 2020)

2. For assembled and annotated genomes we suggest deposition in one or more of these archives:

    2.1. NCBI GenBank (in FAIRsharing), accessible through NCBI Virus (Hatcher EL, 2017; in FAIRsharing), for submission documentation see Viral sequence submission documentation (NIH-NCBI, 2020)

    2.2. DDBJ Annotated/Assembled Sequences (DDBJ, 2020)

    2.3. ENA (EMBL-EBI)

3. Virus Data submitted to GenBank (NCBI) and RefSeq (NCBI) will be available for re-use through NCBI Virus (NCBI)

4. 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.

5. Before submission of raw sequence data (e.g., shotgun sequencing) to INSDC archives, it is necessary to remove contaminating human reads.

6.4.1.2 Data and Metadata Standards

A list of relevant genomics data and metadata standards can be found in FAIRsharing (FAIRsharing, 2020), some specific examples are:

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 (Cock PJA, 2009; in FAIRsharing); optionally add compression with gzip

1.2. Genome contigs: .fastq (Cock PJA, 2009; in FAIRsharing); if uncertainties of the assembler can be captured, .fasta (Pearson WR, 1988; in FAIRsharing) otherwise; optionally add compression with gzip

1.3. De novo aligned sequences: .afa

1.4. Gene Structure: .gtf (in FAIRsharing)

1.5. Gene Features: .gff (in FAIRsharing)

1.6. Sequences mapped to a genome: .sam (Li H, 2009; in FAIRsharing) or the compressed formats .bam (in FAIRsharing) or .cram (Fritz MH, 2011). 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.

1.7. Variant calling: .vcf (in FAIRsharing). 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.

1.8. Browser: .bed (in FAIRsharing)

2. Consider annotating virus genomes using the ENA virus pathogen reporting standard checklist (European Nucleotide Archive, 2020), which is a minimal information standard under development right now and the more general Viral Genome Annotation System (VGAS) (Zhang K, 2019).

3. For submitting data and metadata relating to phylogenetic relationships (including topology, branch lengths, and support values) consider using widely accepted formats such as:

3.1. Newick (Felsenstein, 1986; in FAIRsharing)

3.2. NEXUS (Maddison, 1997; in FAIRsharing)

3.3. PhyloXML (Han & Zmasek, 2009; Stoltzfus, 2012; in FAIRsharing)

3.4. The Minimum Information About a Phylogenetic Analysis checklist provides a reference list of useful tree annotations (Lapp H et al., 2017; in FAIRsharing).

6.4.2 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 Chapter 9 on Ethical and Legal Compliance and not repeated here. The COVID-19 host genetics initiative is a bottom-up collaborative effort to generate, share, and analyze data to learn the genetic determinants of COVID-19 susceptibility, severity, and outcomes.

6.4.2.1 Generic Recommendations

1. 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.
2. Especially for raw sequencing data make sure to include Quality Control (QC) results and details of the sequencing platform used.
3. Common terminologies for reporting statistical tests, e.g., with StatO (in FAIRsharing), enable reuse and reproducibility.
4. Researchers interested in human leukocyte antigen (HLA) genomics are referred to the HLA COVID-19 consortium.

6.4.2.2 Repositories[1]

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

1. **Gene expression data** should in general be retrieved from or deposited in the repositories listed below (Blaxter M et al., 2016). To achieve load balancing, it is recommended to choose the respective regional repository. It should be noted that INSDC resources (i.e. DDBJ, ENA and NCBI) synchronize most of their data sets daily[2].

   1.1 Transcriptomics of human subjects (requiring authorized access):

   1.1.1 Database of Genotypes and Phenotypes (dbGaP) (Mailman MD et al., 2007; in FAIRsharing)

   1.1.2 European Genome-Phenome Archive (EGA) (Lappalainen I et al., 2015; in FAIRsharing). The corresponding non-sensitive metadata will be available through EBI ArrayExpress (Athar A et al., 2019; in FAIRsharing).

   1.1.3 Japanese Genotype-phenotype Archive (JGA) (Kodama Y et al., 2015; in FAIRsharing)

   1.2 Transcriptomics (from cell lines/animals):

   1.2.1 ArrayExpress (Athar A et al., 2019; in FAIRsharing)

   1.2.2 Gene Expression Omnibus (Barrett T et al., 2013; in FAIRsharing)

   1.2.3 Genomic Expression Archive (in FAIRsharing)

   1.3 Underlying reads can be retrieved from/will automatically deposited to the corresponding read archive:

   1.3.1 DDBJ Sequence Read Archive (DRA) (Kodama Y et al., 2012; in FAIRsharing), for submission documentation see here

   1.3.2 European Nucleotide Archive (in FAIRsharing), for submission documentation see here

   1.3.3 NCBI Sequence Read Archive (SRA) (in FAIRsharing), for submission documentation see here

[1] The lists of repositories here are sorted alphabetically within each section. The order should not be interpreted as any kind of preference of recommendation

[2] This does not include the sections for restricted access data (dbGaP, EGA, JGA) and for gene expression (ArrayExpress/GEA/GEO)

1.4 Microarray-based gene expression data:
  - 1.4.1 ArrayExpress (Athar A et al., 2019; in FAIRsharing)
  - 1.4.2 Gene Expression Omnibus (Barrett T et al., 2013; in FAIRsharing)
  - 1.4.3 Genomic Expression Archive (in FAIRsharing)

1.5 Data on the originating sample can be retrieved from/will automatically deposited to the corresponding sample archive:
  - 1.5.1 DDBJ BioSample
  - 1.5.2 EBI BioSamples (in FAIRsharing)
  - 1.5.3 NCBI BioSample (in FAIRsharing)

1.6 For specialized use cases, additional domain-specific repositories might exist, a curated list of which can be found in FAIRsharing. Data depositors are encouraged to submit their data to these specialized resources in addition to one of the resources mentioned above.

2. **Genome-Wide Association Studies** (GWAS):
   - 2.1 GWAS Catalog (in FAIRsharing)
   - 2.2 EGA (Lappalainen I et al., 2015; in FAIRsharing)
   - 2.3 GWAS Central (in FAIRsharing)

3. **Adaptive Immune Receptor Repertoire Sequencing** (AIRR-seq)[3] data: It is recommended that data be deposited using AIRR Community compliant processes and standards, in either of the following repositories.
   - 3.1 AIRR-seq specific repositories that are part of the AIRR Data Commons, for example the iReceptor Public Archive (Corrie BD et al., 2018) or VDJServer (Christley S et al., 2018; in FAIRsharing).
   - 3.2 INSDC repositories via NCBI SRA/Genbank, following the AIRR Community recommended NCBI submission processes.

6.4.2.3 Data and Metadata Standards

1. **Gene Expression Data**

1.1. Transcriptomics
  - 1.1.1 Preferred minimal metadata standard MINSEQE (in FAIRsharing)
  - 1.1.2 Preferred file formats (sequencing-based):
    - Raw sequences: .fastq (Cock PJA et al., 2010; in FAIRsharing), optional compression with gzip or bzip2
    - Mapped sequences: .sam (in FAIRsharing), compression with .bam (in FAIRsharing) or .cram (Fritz MHY et al., 2011)
    - Transcripts per million (TPM): .csv
  - 1.1.3 Also see FAIRsharing using the query 'transcriptomics'

1.2. Microarray-based gene expression data
  - 1.2.1 Preferred minimal metadata standard: MIAME (Brazma A et al., 2001; in FAIRsharing)
  - 1.2.2 Preferred file formats: tab-delimited text, raw data file formats from commercial microarray platforms (Annotare accepted formats; Athar A et al., 2019)

[3] 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.

2. Genome-wide association studies (GWAS):
    2.1. Preferred minimal metadata standard: MIxS (Yilmaz P et al., 2011; in FAIRsharing)
    2.2. Preferred file formats:
        2.2.1 Binary files: .bim, .fam and .bed (Chang CC et al., 2015; in FAIRsharing)
        2.2.2 Text-format files: .ped and .map (Chang CC et al., 2015)
3. Adaptive Immune Receptor Repertoire sequencing (AIRR-seq):
    3.1. Preferred minimal metadata standards: MiAIRR (Rubelt F et al., 2017; in FAIRsharing)
    3.2. Preferred file formats:
        3.2.1 AIRR repertoire metadata, formatted as .json or .yaml (Vander Heiden JA et al., 2018)
        3.2.2 AIRR rearrangements, formatted as .tsv (Vander Heiden JA et al., 2018; in FAIRsharing)

6.4.3 Recommendations for Structural Data

6.4.3.1 Repositories

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

1. Structural data on proteins acquired using any experimental technique should be deposited in the wwPDB: Worldwide Protein Data Bank (Burley SK et al., 2019; in FAIRsharing) ; a collaborating cluster of three regional centers at (1) Europe EBI PDBe (PDBe-KB consortium, 2020; in FAIRsharing) and The Electron Microscopy Data Bank EMDB (Lawson CL et al., 2011; in FAIRsharing) (2) USA RCSB PDB (Berman HM, 2000 in FAIRsharing and (3) Japan PDBj (Kinjo AR et al., 2017; in FAIRsharing). Data submitted to either of these resources will be available through each of them.

2. A public information sharing portal and data repository for the drug discovery community, initiated by the Global Health Drug Discovery Institute of China (GHDDI) is the GHDDI Info Sharing Portal and includes the following:
    2.1. compound libraries including the ReFRAME compound library (Janes J et al., 2018)(the world's largest collection of its kind, containing over 12,000 known drugs), a diversity-based synthetic compound library, a natural product library, a traditional Chinese medicine extract library
    2.2. Drug Discovery Cloud Computing System on Alibaba Cloud
    2.3. Data mining and integration of historical drug discovery efforts against coronavirus (e.g. SARS/MERS) using AI and big data
    2.4. Molecular chemical modeling and simulation data using computational tools.

6.4.3.2 Locating Existing Data

1. The COVID-19 Molecular Structure and Therapeutics Hub community data repository and curation service for structure, models, therapeutics, simulations and related computations for research into the SARS-CoV-2 / COVID-19 pandemic is maintained by The Molecular Sciences Software Institute (MolSSI) and BioExcel.

6.4.3.3 Data and Metadata Standards

1. X-ray diffraction

1.1. There are no widely accepted standards for X-ray raw data files. Generally these are stored and archived in the Vendor’s native formats. Metadata is stored in CBF/imgCIF format (in FAIRsharing) (See: catalogue of metadata resources for crystallographic applications)

1.2. Processed structural information is submitted to structural databases in the PDBx/mmCIF format. (Fitzgerald PMD, 2006; in FAIRsharing)

2. Electron microscopy

2.1. Data archiving and validation standards for cryo-EM maps and models are coordinated internationally by EMDataResource (EMDR).

2.2. Cryo-EM structures (map, experimental metadata, and optionally coordinate model) is deposited and processed through the wwPDB OneDep system (wwPDB Consortium. https://deposit-2.wwpdb.org/), following the same annotation and validation workflow also used for X-ray crystallography and NMR structures. EMDB holds all workflow metadata while PDB holds a subset of the metadata.

2.3. Most electron microscopy data is stored in either raw data formats (binary, bitmap images, tiff, etc.) or proprietary formats developed by vendors (dm3, emispec, etc.).

2.4. Processed structural information is submitted to structural resources as PDBx/mmCIF (Fitzgerald PMD et al., 2006; wwPDB, 2014; in FAIRsharing).

2.5. Experimental metadata include information about the sample, specimen preparation, imaging, image processing, symmetry, reconstruction method, resolution and resolution method, as well as a description of the modeling/fitting procedures used and are described in EMDR, see also Lawson et al (Lawson C et al., 2020).

3. NMR

3.1. There are no widely accepted standards for NMR raw data files. Generally these are stored and archived in single FID/SER files.

3.2. One effort for the standardization of NMR parameters extracted from 1D and 2D spectra of organic compounds to the proposed chemical structure is the NMReDATA initiative and the NMReDATA format (DOI: 10.1002/mrc.4737). The goal of the NMReDATA initiative is to improve the FAIRness and quality of the NMR data. The NMReDATA format allows for data organization in a way that the assignment data can be stored in a reliable manner (using DOI) and allowing for their verification against the experimental spectra.

3.3. There is no universally accepted format, especially for crucial FID-associated metadata. NMR-STAR (Ulrich et al, 2018; in FAIRsharing) and its NMR-STAR Dictionary (Ulrich & Wedell, 2019) is the archival format used by the Biological Nuclear Magnetic Resonance data Bank (BMRB) (in FAIRsharing), the international repository of biomolecular NMR data and an archive of the Worldwide Protein Data Bank (Burley, SK 2019 in FAIRsharing).

3.4. The nmrML format specification (XML Schema Definition (XSD) and an accompanying controlled vocabulary called nmrCV) are an open mark up language and an ontology for NMR data (PhenoMeNal H2020 project, 2019; in FAIRsharing).

3.5. Processed structural information is submitted to structural databases in the PDBx/mmCIF format (Fitzgerald PMD et al., 2006; wwPDB, 2014; in FAIRsharing).

4. Neutron scattering

4.1. ENDF/B-VI of Cross-Section Evaluation Working Group (CSEWG) and JEFF of OECD/NEA have been widely utilized in the nuclear community. The latest versions of the two nuclear reaction data libraries are JEFF-3.3 (Cabellos O, 2017) and ENDF/B-VIII.0 (Brown DA, 2018) with a significant upgrade in data for a number of nuclides (Carlson AD, 2018).

4.2. Neutron scattering data are stored in the internationally-adopted ENDF-6 format (Herman and Trkov, 2010) maintained by CSEWG.

4.3. Processed structural information is submitted to structural databases in the PDBx/mmCIF format (Fitzgerald PMD et al., 2006; wwPDB, 2014; in FAIRsharing).

5. Molecular Dynamics (MD) simulations (e.g., of SARS-CoV-2 proteins)

5.1. Raw trajectory files containing all the coordinates, velocities, forces and energies of the simulation are stored as binary files: .trr, .dcd, .xtc and .netCDF; See also a description of metadata standards to be considered.

5.2. Refined structural models from experimental structural data using MD simulations are stored in .pdb format (Bernstein FC et al., 1977).

6. Computer-aided drug design data

6.1. Virtual screening results are stored in 3D chemical data formats such as:

6.1.1. .mol, .sdf and .mol2 (Hoffman, 2020)

6.1.2. .pdb (Bernstein FC et al., 1977)

6.1.3. Structural formulas either in SMILES (Anderson E, Veith GD, Weininger D, 1987; in FAIRsharing)

6.1.4. IUPAC International Chemical Identifier (InChI), and identified through InChIKey, a non-proprietary identifier for chemical substances that can be used in printed and electronic data sources thus enabling easier linking of diverse data compilations (Heller SR et al., 2015; in FAIRsharing).

6.4.4 Recommendations for Proteomics

Proteomics studies are used to find biomarkers for disease and susceptibility.

6.4.4.1 Repositories

1. For a curated list of relevant repositories see FAIRsharing using the query 'proteomics'. The ProteomeXchange Consortium (in FAIRsharing) enables searches across the following deposition databases, following common standards

1.1. For shotgun proteomics one of:

1.1. PRIDE (Perez-Riverol Y et al, 2019; FAIRsharing)

1.2. MassIVE (Wang M et al, 2018; in FAIRsharing)

1.3. jPOST | Japan Proteome Standard Repository/Database (Okuda S et al, 2017; in FAIRsharing)

1.4. iProX - integrated Proteome resources (Ma J, 2019; in FAIRsharing)

1.2. For targeted proteomics one of:

1.5. PASSEL (Farrah T et al, 2012; Kusebauch U et al, 2014; in FAIRsharing)

1.6. Panorama (Sharma V et al, 2018; Sharma V et al, 2014)

1.3. For repossessed results one of:

1.7. PeptideAtlas (Deutsch EW, 2009; in FAIRSharing)

1.8. MassIVE (Wang M et al, 2018; in FAIRsharing)

2. Non-mass spectrometry based protein oriented data (ELISA, Luminex, ELISPOT, neutralizing antibody titer), Flow Cytometry, Mass Cytometry and HLA/KIR typing data can be submitted to ImmPort (Bhattacharya S et al. 2018; in FAIRsharing).

6.4.4.2 Data and Metadata Standards

1. For a curated list of relevant standards see FAIRsharing using the query 'proteomics'. Specific examples:

   1.1. Use the minimal information model specified in MIAPE (HUPO Proteomics Standards Initiative, 2007 & Taylor CF et al, 2007; in FAIRsharing) and these are filled using the controlled vocabularies specified by the Proteomics Standards Initiative, PSI CVs (FAIRsharing)

   1.2. Recommended formats are:

      1.2.1. gelML (in FAIRsharing)

      1.2.2. TraML (HUPO Proteomics Standards Initiative, 2017; in FAIRsharing)

      1.2.3. mzML (HUPO Proteomics Standards Initiative, 2017; in FAIRsharing)

      1.2.4. mzTab (HUPO Proteomics Standards Initiative, 2017; in FAIRsharing)

      1.2.5. For protein quantization data: mzQuantML (HUPO Proteomics Standards Initiative, 2017; in FAIRsharing),

      1.2.6. For protein identification data: mzIdentML (HUPO Proteomics Standards Initiative, 2017; in FAIRsharing)

2. Flow Cytometry

   2.1. Use the minimal information model MiFlowCyt (Lee JA et al., 2008; in FAIRsharing)

   2.2. Recommended formats are:

      2.2.1. .fcs (Spidlen J et al., 2010; in FAIRsharing)

      2.2.2. Gating-ML (Spidlen J et al., 2015; in FAIRsharing)

6.4.5 Recommendations for Metabolomics

Metabolomics studies are used to find biomarkers for disease and susceptibility. Lipidomics is a special form of metabolomics, but is also described in more detail in a separate section below because of its special relevance to COVID-19 research.

6.4.5.1 Repositories

1. For a curated list of relevant repositories see FAIRsharing using the query 'metabolomics'.

2. Metabolomics data can be submitted to:

   2.1. MetaboLights (in Europe) (Haug K et al., 2020; in FAIRsharing)

   2.2. Metabolomics Workbench (in the USA) (Sud M et al., 2016; in FAIRsharing).

6.4.5.2 Data and Metadata Standards

3. For a curated list of relevant standards see FAIRsharing using the query 'metabolomics'. Specific examples:

   3.1. Core Information for Metabolomics Reporting, CIMR standard (in FAIRsharing)

   3.2. For identifying chemical compounds use SMILES (Anderson E, Veith GD, Weininger D, 1987; in FAIRsharing) or InChI (Heller SR et al., 2015; in FAIRsharing).

3.3. To document Investigation/Study/Assay data, use the ISA Abstract Model, also implemented as a tabular format, ISA-Tab in MetaboLights (in Europe) (Haug K et al., 2020; in FAIRsharing) and Metabolomics Workbench (USA) (Sud M et al. 2016; in FAIRsharing). For an introduction to ISA, please read Towards interoperable bioscience data (Sansone S-A et al, 2012).

3.4. Recommended formats are:

3.4.1. For LC-MS data use: ANDI-MS specification (ASTM E1947), an analytical data interchange protocol for chromatographic data representation and/or mzML (HUPO Proteomics Standards Initiative, 2017; in FAIRsharing).

3.4.2. For NMR data: nmrCV (in FAIRsharing), nmrML (PhenoMeNal H2020 project, 2019; in FAIRsharing).

6.4.6 Recommendations for Lipidomics

Lipidomics revealed an altered lipid composition in infected cells and serum lipid levels in patients with preexisting conditions. Lipid rafts (lipid microdomains) play a critical role in viral infections facilitating virus entry, replication, assembly and budding. Lipid rafts are enriched in glycosphingolipids, sphingomyelin and cholesterol. It is likely that coronavirus (SARS-CoV-2) enters the cell via angiotensin-converting enzyme-2 (ACE2) that depends on the integrity of lipid rafts in the infected cell membrane.

6.4.6.1 Generic Recommendations for Researchers

Lipidomics analysis should follow the guidelines of the Lipidomic Standards Initiative

6.4.6.2 Repositories

The largest repository for lipidomics data is MetaboLights (Haug K et al., 2020; in FAIRsharing)

6.4.6.3 Data and Metadata Standards

1. Metadata standards

1.1. Metadata should follow recommendations from the CIMR standard by the Metabolomics Standards Initiative (FAIRsharing). It should be made available as tab or comma separated files (.tsv or .csv).

2. Data standards

2.1. Data can be stored in LC-MS file, tabular (.tsv) or comma (.csv) formats :

3. Data analysis

3.1. Most of the analysis is usually performed using the software delivered by the suppliers of the instrumentation. In line with generic software recommendations it should be made sure that the process and parameters are well described, and that the output is converted to a standard format.

3.2. Workflow for Metabolomics (W4M) is a collaborative portal dedicated to metabolomics data processing, analysis and annotation for Metabolomics community

3.3. Data processing using R software and associated packages from Bioconductor (xcms, camera, mixOmics) is a flexible and reproducible way for lipidomic data analysis

4. Compound identification

4.1. After data processing, potential biomarkers should be annotated. This could be done either by manual (Lipid Maps tools) or automated identification against templates (Library templates for compounds identification) with the help of softwares such as LipidBlast, MSPepSearch or MS-DIAL. Finally, lipids classification and nomenclature should follow the LIPID MAPS guidelines

7. Data Sharing in Social Sciences

7.1 Focus and Description

Data from the social sciences is essential for all domains (including omics, clinical and epidemiology, among others) that seek to better plan for effective management of COVID-19 pandemic and its consequences. Social scientists are collecting new information and reusing existing data sources to better inform leaders and policymakers about pressing social and economic issues regarding COVID-19, to enable evidence-based decision-making, as this pandemic is as much a social as it is a medical phenomenon. Social science research, involving a predominance of observational methods, produces unique data that cannot be recreated in the future. Furthermore, key social science data, such as demographics, are valuable tools for all disciplines to be able to understand context and link datasets. 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 research data management is expedient--but not hasty--and that data contributions from the social sciences are shared and preserved in ways that allow them to be leveraged long-term for the broadest impact and reused across all domains.

7.2 Scope

Social science disciplines include economics, sociology, political science, education, demography, social anthropology, geography, and psychology, among others. The current health crisis is influenced by the way political leaders, health expert panels, social communities and individual citizens have reacted to the challenges presented by the virus. Social science data have significant value for tracking and altering the social, political, cultural, psychological and economic impact of COVID-19 as well as future health emergencies. Such knowledge can facilitate preparation and mitigation measures, ameliorate negative impacts, improve social and economic wellbeing, and inform decision-making processes. These recommendations are shaped by the need for rapid and long-term access to social science data in the following areas, among others:

1. Social Isolation and Social Distancing,
2. Family and Intergenerational Relationships,
3. Quality of Life and Wellbeing,
4. Health Behaviors and Behavior Change,
5. Health Disparities,
6. Impact on Vulnerable Populations (including immigrants, minority groups),
7. Community Impact and Neighbourhood Effects,
8. Transportation; Food Security,
9. Beliefs, Attitudes, Misinformation, Public Opinion,
10. Technology-Mediated Communication (public information campaigns; social media use),
11. Economic Impacts (including industry, work, unemployment),
12. Organisational Change,
13. Social Inequalities and Discrimination,
14. Education Impacts (including online learning),
15. Political Dynamics, Policy Approaches, and Government Expenditure,
16. Criminal Justice (including domestic violence, prison populations, cybercrime),
17. Human Mobility and Migration (including dislocation).

Ensuring data produced across such areas of research are readily accessible and properly documented will (1) advance the social science research agenda around COVID-19; (2) promote interoperable cross-disciplinary and cross-cultural data use, collaboration, and understanding and (3) build a foundation for managing social science data during pandemics and health emergencies more generally, ensuring that social science research can be leveraged for the public good.

7.3 Policy Recommendations

In formulating policies on pressing questions in times of emergency, policymakers require access to social science research based on data. Because the COVID-19 crisis is taking place in the context of a data-intensive economy, data play a crucial role and has value across many stakeholders (e.g., scientists, citizens, governments, private corporations). Data generated from public funded projects should be made quickly available to the research community. The more merit users access the data the more knowledge about COVID-19 is generated. The following recommendations are aimed at ensuring the policies and practices across the wide array of organizations supporting research during COVID-19 require and ensure high quality, social science data in line with FAIR principles.

1. Ensure robust funding streams for social science research, which is essential to the work in all other research domains and important itself for understanding and managing the social, behavioural, and economic aspects of pandemics. This is necessary to avoid increasing health and social disparities due to COVID-19 and other health emergencies.
2. Funding decisions should prioritize projects where the social science data being produced can be used across domains and are linkable and interoperable.
3. Social science funding should both require data sharing and support infrastructure for data archiving and preservation. This includes striving for funding models that are applied equitably across projects, researchers, and countries. This is also a mandate for covering costs for infrastructure in the broadest sense (e.g., ensuring open access to data, curation services, research data management costs across the lifecycle, and long-term preservation, among others).
4. As funding agencies make use of rapid funding mechanisms (e.g., administrative supplements, fast-track projects), it should not be at the expense of requiring Data Management Plans and ensuring data are sharable.
5. Big tech companies hold data that can help understand the pandemic better. Data sharing policies should facilitate data flows from data holders to the scientific community with the goal of protecting citizens' rights and health (Askitas 2018).
6. Within research institutions & funding agencies, ensure structures for researchers to get credit and support for publishing data as valid research outputs, which can help researchers to invest sufficient time in this work. This can include developing research assessment systems that reward data outputs, alongside publications and other research objects.
7. Despite rapid needs for data and research, basic human subject protections must be upheld by all institutions engaged in research. All human subjects are equal and should be treated as such, every single human subject should be treated fairly.
8. All stakeholders (researchers, research institutions, institutional ethics review boards/ethics committees, healthcare organizations, funding agencies and policy makers) should consider COVID-19 data sharing needs while reviewing the ethics standards, finding balance between the community good and the individual rights of the participants.

9. Even in disciplines where there is less tradition for data sharing, the pandemic offers an opportunity for funding organizations to advance data sharing practices by requiring data sharing universally across research projects where possible.
10. Research institutions should provide researchers with robust and secure data storage facilities that follow recommendations regarding areas such as regular backup in multiple locations and data protection.
11. 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.
12. Official statistical agencies should ensure that there are uniform recommendations about the minimal number of metadata variables shared that will allow linking the different types of data produced around COVID-19 (e.g., geospatial codes and time stamping using controlled vocabularies, ideally international standards such as NUTS (eurostat, n.d.) and ISO (ISO, n.d.)).
13. Repositories handling social science data should make metadata available under a Creative Commons Public Domain Dedication (CC 0 1.0) or equivalent, in line with the FAIR principles (in particular machine-actionable).
14. Social science journals should require COVID-19 related articles to provide data statements on data availability that point to access in a publicly available repository.

7.4 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. The following recommendations in relation to data management and sharing, ethical and legal issues, metadata, storage, should be referenced in making decisions which necessarily balance individual and public rights and benefits.

7.4.1 Data Management Responsibilities and Resources

To ensure broad reuse, a Data Management Plan (DMP) constructed for social science data collections should guide the handling of the data over time and help all disciplines (e.g., clinical, epidemiology) understand the data.

Researchers should create a DMP at the beginning of the research process so that it can be included in the work plan and the budget. The DMP is a “living” document, which may change over the course of a project.

Projects already underway that might contribute data to address COVID-19 should update their DMPs to ensure alignment with current recommendations.

Consider resources required (including the time of research staff to manage data), balanced against the costs of not doing this properly. When writing a grant proposal, check the guidelines of the funder with regard to the eligibility of data management costs and, if applicable, include them in the budget.

All parties with responsibility for activities across the research lifecycle - not just the researcher - have a part to play in ensuring good quality data that is safeguarded so it can be located, understood, and effectively used and reused. Roles and the responsibilities should be considered early (ideally at the data planning phase), and be clearly defined and documented in the DMP. A common understanding of how data will be managed is particularly important in collaborative projects that involve many researchers, institutions and groups with different ways of working.

When writing the DMP, researchers should contact institutional support services (e.g., library staff), the repository of their choice, or other research infrastructure providers which may offer guidelines for the DMPs in advance of deposit.

Consult a list of data management resources in Adapt your Data Management Plan (CESSDA, 2019) and associated DMP template (CESSDA, 2018). Use one of the DMP tools for your country, funder, and preferred language: DMPonline (DCC, n.d.); DMPTool (University of California, n.d.); DMP assistant (Portage, n.d.); ARGOS (OpenAIRE, n.d.); DMP OPIDoR (OPIDoR, n.d.) and Plan de Gestión de Datos (PGD) (Vilches, n.d.) and address the relevant aspects of making the data FAIR (Wilkinson et al., 2016) in a DMP.

Researchers should aim to register their DMP as an openly accessible, public deliverable.

7.4.2 Documentation, Standards, and Data Quality

Social science data producers should provide thorough documentation about the data themselves, the research context, methods used to collect, store, and treat data, and quality-assurance steps taken. Consider the needs of the future data user when developing and creating documentation. The documentation serves multiple purposes, supporting reproducibility, linkage, quality checking, understandability and transparency of the collection and storage process.

Documentation for data elements (e.g., geography, time period, demographics) that are useful for linking to other sources of data around COVID-19 should allow full understanding of context, method, and limitations),

Use standardized codes for places to reduce data consistency challenges that come from the use of textual entity names. We strongly encourage the use of ISO-3166 for countries or administrative subdivisions, ANSI and FIPS for U.S. States and counties, and standard identifiers for organisational entities such as companies (Coffey, n.d.; INSEAD, n.d.). This set of actions will facilitate data analysis, harmonization, linking, visualization, and integration in applications.

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. For example, the use of readme files to orient a user to a set files are common in some professions.

To ensure that data is more generalizable, researchers should provide access to information that can be used to address selection bias (e.g., demographic characteristics, geographic variables, etc.). Furthermore, data should be collected in a standardized way at least at the national level to enable cross-country comparison. COVID-19 has offered important insights into the variation amongst how countries experience and respond to pandemics. Effective cross-national comparisons can provide useful insights for the development of future global emergency preparedness programmes.

Data should be stored in at least one non-proprietary format that is well-documented. Many repositories publish lists of recommended, preferred, or acceptable formats that are useful to consult. Two sources for recommended formats are UKDA Recommended Formats (UK Data Service, n.d.) and Library of Congress Recommended Formats Statement (Library of

Congress, n.d.). The social sciences benefits from the use of many common formats used across disciplines, enabling broader interoperability.

7.4.3 Storage and Backup

Where possible, researchers should use the official storage provisions available from their institution, rather than personal storage, including when working remotely, as they are more likely to provide robust backup and data protection features.

Researchers with sensitive data or data with disclosure risk should seek a storage solution for their data requiring remote access to safeguard data which offers flexibility and protection (German Data Forum (RatSWD), 2020).

Social sciences data, as is true for human subjects data in other domains, 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 that often vary by country.

Data access while data collection is active should be limited to those with authorisation to use the data. To speed access to COVID-related data, we encourage authorizing external user groups where possible. 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, and no matter where or how it is stored.

Where possible, best practice is to store data (including participant consent files) without direct identifiers and replace personal identifiers with a randomly assigned identifier. Researchers should create 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.

Ensure that data should be backed up in multiple locations all under the same security conditions.

Where possible, select a storage solution that allows an easy way to maintain version control.

7.4.4 Legal and Ethical Requirements

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. 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.

Ethics review during a crisis like the COVID-19 pandemic is critical to protect highly vulnerable populations from potential harm. Therefore these Guidelines endorse 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 (AAS, 2020).

Respect Indigenous Peoples rights and interests and follow the CARE Principles for Indigenous Data Governance (Research Data Alliance International Indigenous Data Sovereignty Interest Group, 2019), that complement the FAIR principles and are people and purpose-oriented.

Ethical use of open data ensures inclusive development and equitable outcomes. Metadata should acknowledge the provenance and purpose and any limitations or obligations in secondary use inclusive of issues of consent.

Researchers whose data have legal, privacy, or other restrictions should seek out appropriate alternative avenues for data sharing including restricted access conditions and embargoes, only when absolutely essential.

Ensure licenses and agreements in data acquisition enable downstream data sharing and preservation. The way primary data have previously been collected and processed may have an impact on the sharing and use of secondary data. Sharing and use of these data can be agreed for a certain duration, defined purposes and with appropriate guarantees for both researchers and data providers. License for secondary data (e.g., with universities or research groups) should be written to allow researchers to share data, to enable broader sharing for the public good, such as limited extracts that cannot undermine the data provider's business model. Researchers should seek local support to clarify how best to share secondary data, to ensure and negotiate the appropriate rights.

If you work with commercial partners, seek opportunities to negotiate data sharing mechanisms agreeable to both parties. Develop partner and consortial agreements that make explicit each partner's rights, including what data can be shared and how. Ensure equitable partnerships.

Using data from social media introduces additional issues. Individuals creating and sharing content may not regard this as a public space and have an expectation of a degree of privacy. Furthermore, social networks by definition reveal connections between many individuals; thus an individual post or *tweet* may provide information on many different data subjects without their knowledge or consent. In addition, researchers collecting data from the web should ensure they have sufficient rights to do so to safeguard their ability to use the data; many websites have terms and conditions that prohibit data collection, particularly via web scraping and other automated methods.

7.4.5 Data Sharing and Long-term Preservation

Disciplinary norms vary widely across the multiple social sciences disciplines in relation to how common it is for data to be shared and deposited. Some disciplines, including political science and economics, have rapidly developed data sharing practices based on widely shared norms about the replicability and transparency of research findings, as well as pre-registration of research studies. These have often been fostered by the requirements of top international journals to make data available for validation. Adoption levels vary considerably across countries even within disciplines, mostly as a function of the requirements and compliance monitoring of funders.

Embracing the FAIR agenda is now critical for all social scientists collecting data relating to COVID-19, and future pandemics, in order to ensure maximum benefit from the data. In the current emergency context, it is a moral imperative to preserve the data and share it in the most open way possible for each case.

Where possible, provide immediate open access to all relevant research data. Open data should be licensed under Creative Commons Attribution 4.0 International License (Creative Commons, n.d.-a) or a Creative Commons Public Domain Dedication (Creative Commons, n.d-b.) 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.

Deposit quality-controlled research data in a data repository, whenever possible in a trustworthy digital repository committed to preservation. As the first choice, disciplinary repositories are recommended for maximum visibility, followed by general or institutional repositories. Furthermore, these Guidelines encourage depositing in repositories that: have undergone formal certification (such as Core Trust Seal (CoreTrustSeal, n.d.), nestor Seal for Trustworthy Digital Archives (nestor, n.d.) or ISO 16363 (PTAB, n.d.)); are included in re3data.org (DataCite, 2020) and the RDA-Endorsed FAIRsharing (FAIRsharing, n.d.); are indexed in DataCite; and are part of Research Infrastructures (e.g., CESSDA, CLARIN and others), as this also ensures maximum cross-border visibility. Aim for repositories that comply better with the FAIR principles. 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.

If you use a general repository (e.g. Figshare, Dataverse, Dryad, openICPSR, Zenodo and others), describe the data using the following at a minimum: the dataset's creator, title, description, year of publication, any embargo, licensing terms, identifier, etc.

To ensure social sciences data can be linked with data being produced by other entities, consider long-term preservation of information that enables data linkages to be made over time, 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.

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 it with the data where possible.
Researchers should deposit in a repository data that underpin published findings, data that allow for validation and replication of results, and the broader set of data with long-term value.

8. Research Software and Data Sharing

8.1 Focus and Description

It is important to put forward some key practices for the development and (re)use of research software, as these facilitate sharing and accelerate results in responses to the COVID-19 pandemic.

A number of foundational, clear and practical recommendations around research software principles and practices are provided here, in order to facilitate the open collaborations that can contribute to addressing the current challenging circumstances. 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 (Akhmerov et al., 2019; Clément-Fontaine et al., 2019; Jiménez et al., 2017; Lamprecht et al., 2019; Wilson et al., 2017).

8.2 Scope

These recommendations cover general practices, not details of particular technologies or software development tools. The recommendations in Section 8.5 (Guidelines for researchers) will help researchers improve the quality and reproducibility of their software but should also have an impact on policy makers, funders and publishers. The aim is that researchers follow the principles generally, even if they cannot follow one or more of the principles completely, because following the principles will improve the research environment for themselves and others. With the recommendations in Section 8.3 (Policy recommendations), we aim for policy makers and funders to realise the, sometimes behind the scenes, work around research software, e.g., documentation and maintenance, so they create opportunities addressing, for instance, the acquisition of skills and full development cycle. With the recommendations in Section 8.4 (Guidelines for publishers), we aim for publishers to push forward to citable software so it becomes equal in recognition to data and scholarly publications as a research outcome.

Throughout this document we will be using software as a placeholder and interchangeably for compiled software (i.e., binaries) as well as for software source code. When necessary to differentiate, we will make an explicit comment.

8.3 Policy Recommendations

Research software is essential for research, and this is increasingly recognised globally by researchers. This section provides recommendations for policy makers and funders on how to support the research software community to respond to COVID-19 challenges, based on existing work (Akhmerov et al., 2019). 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. Additionally, given the impact that funding agencies can have in shaping research, it is equally important to ensure that research software is recognised and acknowledged as a direct and measurable outcome of funded efforts.

8.3.1 Support the funding of development and maintenance of critical research software.

Policy makers and funders 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. By providing the resources that are necessary to adhere to best software development practices, policy makers and funders are making it easier for researchers to move from quick and dirty coding to creating shared and reproducible software, allowing implementation of recommendations detailed in Sections 8.5.4 (provide sufficient metadata/documentation) and 8.5.5 ensure portability and reproducibility), thus increasing overall software quality and usefulness. Funding for software development will also enable anyone producing research software to take the time to do it well and to document it, which also aligns with recommendation on Section 8.5.4.

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).

8.3.2 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 code licence, or at least require the software 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 copyright holders. The use of open source code licences should be seen as the default for research software in funded efforts, and policy makers should enact policies to encourage that practice. When software is made available under an open source licence, it means that its underlying source code is made freely accessible, as encouraged by the "A" in FAIR (Findable, Accessible, Interoperable and Reusable) (Wilkinson et al., 2016) 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/or source 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.

The policies and incentives recommended here will motivate researchers to implement recommendations in Sections 8.5.1 (make your software available), 8.5.2 (release your software under a licence) and 8.5.6 (reference your software with Persistent Identifiers) from the good practices section, thus increasing findability, continued usefulness and improvement improvability of software.

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).

8.3.3 Encourage the research community's ability to apply best practices for research software, including training in software development concepts.

Policy makers and funders 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.

The introduction of such programs and funding opportunities will increase the overall understanding and adaptation of all recommendations from the good practices section among researchers, thus supporting the outcomes of the other three recommendations in this section, thus making it easier for researchers to align to all the recommendations provided on Section 8.5 targeting good practices for research software.

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 (Caswell et al., 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).

8.3.4 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 objects. It is equally critical that funders 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. Enacting such policies will encourage researchers to implement recommendations in Sections 8.5.1 (make your software available), 8.5.3 (cite the software you use) and 8.5.6 (reference your software with Persistent Identifiers) from the good practices section, thus creating a self-strengthening system of incentives for the development of high-quality software.

Examples: Policy makers need to support initiatives such as the Declaration on Research Assessment (DORA, 2016), which are beginning to be utilised by research agencies including the Wellcome Trust (Wellcome Trust, 2020), signatories to the Concordat to Support the Career Development of Researchers (Vitae, 2020).

8.4 Guidelines for Publishers

A key component of better research is better software. Publishers can play an important role in changing research culture, and have the ability to make policy changes to facilitate increased recognition of the importance of software in research. This section provides recommendations for publishers on how to support the research software community to respond to COVID-19 challenges.

8.4.1 Require that software citations be included in publications.

It is essential that the role of software in achieving research outcomes is supported. Treating research software as a first class research object in a journal is a very effective mechanism for implementing this as it increases the visibility and credit to the research software developers (for example by enabling academic and commercial citation services and/or databases, such as Google Scholar, Scopus and Microsoft Academic) (Smith et al. 2016).

Examples: The FORCE11 Software Citation Implementation Working Group (Neil et al., 2017) has been leading work in this area for 3+ years. The AAS Journals encourage software citation in several ways (explicit software policy, added the LaTeX \software{} tag to emphasise code used, etc.) (AAS Journals, 2020).

8.4.2 Require that software developed for a publication is deposited in a repository that supports Persistent Identifiers (PIDs).

For publishers to ensure that the research that they publish is reproducible, software developed as part of the work reported in a submission must also be findable. Publishers should require such software to be deposited in a repository that supports PIDs such as Zenodo (CERN, 2020), Figshare (FigShare, 2020) or Software Heritage (Software Heritage, 2020). These repositories provide PIDs that can be directly included in the citation and referenced in a publication, supporting research integrity (Cosmo et al., 2018). If the software is deposited along with data, the selected data repository should provide a PID for the collection. Several versions of the software can be tagged with PIDs and, thus, if multiple versions are used for research, having different PIDs ensures reproducibility.

Example: The Journal of Open Source Software (JOSS) (JOSS, 2020) review process requires authors to make a tagged release of the software after acceptance, and deposit a copy of the repository with a data-archiving service such as Zenodo or figshare. The FORCE11 Software Citation Implementation Working Group (Neil et al., 2017) has been leading work in this area for 3+ years.

8.4.3 Align submission requirements of software publishers to research software best practices.

Research software recently has gained a more prominent place in publishing and some journals specialise in publishing software and software papers. In order to make research software understandable and reusable, it must be produced and maintained using standard practices that follow standard concepts. This can be applied to software ranging from researchers writing small scripts and models; to teams developing large, widely-used platforms. As publishing is an integral part of research, software publishers should enact policies and adopt submission procedures that encourage and support these practices, for example through adopting or adapting software management statements similarly to the widely adopted data management statements.

Example: JOSS requires software to be Open Source and be stored in a repository that can be cloned without registration, is browsable online without registration, has an issue tracker that is readable without registration and permits individuals to create issues/file tickets (JOSS, 2020); SoftwareX submission process includes two mandatory metadata tables that include licence and code availability (ELSEVIER, 2020).

8.5 Guidelines for Researchers

These guidelines aim at supporting researchers with key practices that foster 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 writing scripts or macros.

8.5.1 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 source code itself is shared and freely available (see information about licences 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 (GitHub Inc., 2020b), Bitbucket (Atlassian, 2020), GitLab (GitLab, 2020), etc.).

Resources:
Four Simple Recommendations to Encourage Best Practices in Research Software (Jiménez et al., 2017).
FAIR Software guidelines on code repositories (eScience Center, 2020).

8.5.2 Release your software under a licence.

Software 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 licences typically allow the user to use, study, improve and share your code. You can licence all the software you write, including scripts and macros you develop on proprietary platforms.

Resource: Choose an Open Source License (GitHub Inc., 2020a).

8.5.3 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 setting, this is the most effective way of supporting its continued development and maintenance because it matches the current incentives of that system.

Resource: Software Citation Principles (Smith et al., 2016).

8.5.4 Provide metadata/documentation for others to use your software.

(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.

Resource: Ten simple rules for documenting scientific software (Lee, 2018).

8.5.5 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.

Resource:
Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science (Nüst et al., 2020).
Ten Simple Rules for Reproducible Computational Research (Sandve et al., 2013).

8.5.6 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 (CERN, 2020), Figshare (FigShare, 2020) or Software Heritage (Software Heritage, 2020) which provides more persistent storage than the above code repositories mention in Section 8.5.1. For reproducibility purposes, and if legally allowed, dependencies should also be included in the software deposition.

9. Legal and Ethical Compliance

9.1 Focus and Description

The intention of these guidelines is to help researchers, practitioners and policy-makers deal with all ethical and legal aspects of pandemic response and in particular with regard to key ethical values of equity, utility, efficiency, liberty, reciprocity and solidarity (WHO, 2007; UNESCO, 2020; European Group on Ethics in Science and New Technologies, 2020). In times of public health emergency, it is appropriate to understand how best to consider how best to respond in terms of increased data and research outcome sharing. However, it is important that legal and ethical principles are incorporated into research design from the outset. The law supports research and enables data sharing (EDPB, 2020) . Compliance with the law protects individual researchers, research more generally and the common good. The rule of law cannot be overlooked, therefore, and needs to be taken into consideration along with respect for overarching concerns related to human rights and dignity (Council of Europe, 2020). Especially where marginalisation or other forms of stigmatisation are at stake, these rights and values should inform appropriate research practices directed towards the common good.

These guidelines have been produced by the RDA-COVID working group between March and May 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

9.2 Scope

The COVID-19 pandemic has created significant confusion for researchers in terms of whether, and in which way, existing ethical and legal principles remain relevant. The COVID pandemic does not serve to remove the basic validity of the rights and interests on which these documents and principles are based. In other words, formal protocols for conducting research are required both during a pandemic and at other times, unless otherwise modified by the relevant authorities. 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 time limited adaptations of, or derogation from, normally applicable principles.

The assumption here is that there will be an official statement from WHO of when the international community deems the pandemic to have finished. This may then vary by country. Irrespective of official statement, the necessity and proportionality of any interference with fundamental rights and interests may shift as circumstances change. It will be important to evaluate the continued justification for particular trade-offs at regular intervals in dynamic situations.

A separate, more detailed report expands on the information here

9.3 Policy Recommendations

9.3.1 Initial Recommendations

A set of recommendations are currently being collected and collated. They include:

1. Access to research and research outcomes should be shared with all where possible and in particular, thinking of vulnerable groups and the general focus on solidarity, encouraging the engagement and trust of all participants including vulnerable groups.
2. Ethical guidelines on data collection, analysis, sharing and publication should not be confined to clinical and biological (omic) data. Such guidelines should also extend to all areas of Open Science.
3. In the spirit of the Open COVID Pledge (2020), organisations with potentially useful datasets outside the research communities should be encouraged to make those data available to those research communities during emergency, pandemic situations
4. Ethical and legal policies should be drawn up to monitor and regulate the impact of algorithmic profiling and data analytics, not least in terms of design and implementation.
5. During a pandemic or similar public emergency, ethical review and approval should be expedited, optionally but beneficially involving the public in approval decisions[4]
6. Policy making should be underpinned by empirical research (evidence based) such that decision makers are held to account.
7. Provide guidance and support for non-research organisations to make the data they hold available to the research community.
8. All stakeholders (researchers, policy-makers, editors, funders and so forth) should encourage communication across all disciplines and all areas in the spirit of Open Science.
9. All stakeholders (researchers, editors, funders and so forth) should lobby for regulatory change where existing regulation prevents appropriate data access and sharing.
10. All stakeholders (especially researchers) should be encouraged to publicise practical guidance and advice from their own experience of working through regulatory processes in support of their research.

[4] Cf. the Green / Amber / Red system of risk assessment applied in the UK

In the following version of this document, we also intend to identify areas where more research is needed.

9.3.2 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 (OECD, 2010).
2. The UNESCO International Bioethics Committee (IBC) and World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) in their STATEMENT ON COVID-19: ETHICAL CONSIDERATIONS FROM A GLOBAL PERSPECTIVE (UNESCO, 2020).
3. The Council of Europe points to national resources from national ethics committees or other related to COVID-19: (Council of Europe Bioethics, 2020).
4. The Council of Europe statement on bioethics during COVID-19: (Council of Europe Bioethics, April 2020).
5. The European Group on Ethics in Science and New Technologies statement on solidarity (2020).
6. The Global Alliance for Genomics and Health (GA4GH) Framework for Responsible Sharing of Genomic and Health-Related Data (Global Alliance for Genomics and Health, 2014).
7. The Statement of the African Academy of Sciences' Biospecimens and Data Governance Committee on COVID-19: Ethics, Governance and Community engagement in times of crisis (AAS, 2020).
8. Committee on Economic, Social and Cultural Rights, Statement on the coronavirus disease (COVID-19) pandemic and economic, social and cultural rights (UN Office of the High Commissioner, 2020).
9. RECOMMENDATIONS ON PRIVACY AND DATA PROTECTION IN THE FIGHT AGAINST COVID-19 (Access Now, 2020).

9.4 Guidelines

9.4.1 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 (FORCE11, 2017);
2. The CARE (Collective benefits, Authority to control, Responsibility and Ethics) principles to ensure ethical treatment of individuals and communities (Global Indigenous Data Alliance, 2019);
3. The Global Code of Conduct, specifically Fairness, Respect, Care and Honesty in research activities, to maximise equanimity in research outcome benefit (Schroeder et al, 2020);
4. The Five Safes of research data governance (UK Data Service, 2020; Ritchie, 2008);

5. Research Integrity guidelines (ALLEA, 2017).

9.4.2 Hierarchy of Obligations

Ethics and the law exist in a symbiotic, mutually supportive relationship. Ethical and legal considerations related to research are elaborated in four key types of document: ethical guidelines; policy guidance; codes of conduct; and legal instruments. The distinction between these types of instrument is not always obvious. Regulatory agencies (such as Supervisory Authorities in the EU) do respond to requests for support and clarification. It is therefore recommended that where necessary, researchers work together with the relevant authority to resolve any perceived barriers.

The following principles may prove useful for COVID-19 researchers considering the interaction between instruments:

1. Ethical guidelines are often defined and published by non-law-making bodies, while legal instruments will be adopted by governments or other legislative bodies.
2. Many ethical instruments are *de facto* mandatory for researchers or clinicians, such as those imposed by professional associations or bodies, healthcare institutions, or governmental and funding agencies.
3. Instruments exist in a hierarchy, with legal instruments being generally assumed to take precedence over ethical guidance and policy guidance.
4. Jurisprudence and other official guidelines providing authoritative interpretations of legal instruments will often be complementary to related ethical instruments.In case of dispute, however, the rule of law will prevail.
5. Both legal and ethical instruments should be consulted together to understand all the pertinent issues which need to be taken into consideration.
6. Ethical instruments are generally interpreted harmoniously with the law, and can guide the interpretation of the law if the law does not address a particular issue.

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

1. The obligation to respect confidentiality
2. The obligation to ensure data accuracy
3. The obligation to limit the identifiability of personal data as far as possible - including via pseudonymisation techniques
4. The obligation to use anonymised data instead of personal data, or minimise personal data use, or de-identify where possible
5. The need to process for a specific, authorized, purpose and only to process for secondary purposes provided certain conditions are fulfilled and not processing for purposes beyond scientific research / healthcare e.g. not sharing with employers or other agencies unless mandated by law
6. The obligation to inform individuals about the processing of their data.
7. To hold oneself accountable to, and remain transparent towards, the individuals concerned by the data used

8. To provide individuals access to their data, and to rectify errors or biases in the data on request
9. To allow individuals to object to the processing of their data if required by law
10. To provide individuals the opportunity to request the deletion or return of their data in certain circumstances if this is possible or required by law[5]
11. The obligation to ensure that data are collected from representative sub-populations and not confined to one group[6]
12. The obligation to ensure equal treatment across cohorts to:
    - 12.1. Prevent marginalisation of vulnerable groups
    - 12.2. Encourage engagement from vulnerable groups
    - 12.3. Display trustworthiness and warrant trust (The South African San Institute, 2017)
13. The obligation to share data and the benefits of research outcomes fairly and without regard to discipline, region or country (UNESCO International Bioethics Committee, 2015)
14. The obligation to apply legal and ethical practice to all stages of data collection, processing, analysis, reporting and sharing
15. The obligation for data providers as well as data users to validate and verify the provenance of data, and ensure appropriate consent or other legal basis for the data's use
16. The obligation to ensure that de-identified or aggregated data made public does not contain data elements or rich metadata that could reasonably lead to identify specific persons
17. To validate that data sharing respects the applicable legal requirements, e.g. conclusion of data sharing agreements and/or verifying the legality of a data transfer abroad
18. To consider the legitimacy of the further retention and use of data on persons collected during a public emergency without informed consent, following the emergency.

Such obligations are formalised through ethical guidance (UNESCO, 2005; Council of Europe, 1999, 2010; NHS, 2013). Especially in times of pandemic specific attention to vulnerable groups and guidance on related global justice issues are to be commanded.

9.4.3 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 may find guidance from the following:
    - 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

[5] Some EU Member States, for example, allow for data to be held indefinitely when used for scientific and research purpose

[6] E.g., the traditional white, Caucasian upper-middle-class male.

1.3. the Data Protection Officer will provide support and guidance on data protection issues

1.4. Data and Biospecimen Access Committees will advise on sharing or providing access to data, as well as Intellectual Property issues

1.5. Technology transfer offices provide guidance regarding intellectual property and related issues

1.6. If such bodies are not available at the researcher's home institution the UN Ethics office or national ethics office may be contacted for further support (The United Nations, 2020)

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. Often in cases of health emergencies such as the COVID-19 pandemic fast track procedures are put in place, allowing the approval processes to be accelerated without diminishing the protection of the rights of persons.

9.4.4 Anonymisation

Data will generally be anonymous if they cannot be used to identify a person by all means likely reasonably to be used (Article 29 Working Party on Data Protection, 2007, 2014, 2015). It should be noted, however, that various jurisdictions define the threshold for anonymity differently (for example, the USA). Assessment of all the means reasonably likely to be used must consider not only the data on its own but also the possibility of combination with other accessible data, including by third parties.

The consequence of rendering data anonymous will usually be that certain ethical and legal obligations which usually apply to identifiable data will no longer apply. In particular, anonymisation will usually render data protection law inapplicable. With large datasets, and especially where datasets are cross-correlated, absolute anonymity will often be very hard to achieve. Researchers may thus need to take into account the possibility of future re-identification (see Phillips et al, 2016).

In the European Union, anonymous data falls outside the scope of data protection legislation (GDPR, 2016). A number of tools are available which claim to anonymise personal data, such as sdcMicro (2020) (See also NHS, 2018). However, there are a number of considerations when dealing with data which is said to be anonymous or anonymised. If data are not fully anonymised, then they will fall within the scope of data protection legislation (GDPR, Recital 26) and so therefore require closer controls and management. Check the following recommendations.

1. **De-identified[7] data** can refer to data where personal identifiers have been removed (e.g. US HIPAA). However, there is still some risk that such data may lead to re-identification especially if combined with other data. Generally, de-identification refers to the *process* of reducing data identifiability rather than the identifiability of the resulting data (Phillips and Knoppers, 2016).
2. **Pseudonymised data** are data where personal identifiers have been changed or removed (i.e., personal names and locations obscured). There is a separate key, index, or technological process which links the pseudonymous id code to an individual. The pseudonymisation of data will not reduce the data protection obligations in the data but can be a requirement to the lawful use of data in some jurisdictions and ethical regimes, where practicable (e.g. GDPR).
3. **Data that Cannot be Re-personalised:** Some jurisdictions, such as the EU, recognise a median status for data that remains identifiable by law, but that the controller is not able to reidentify (GDPR art. 11). For instance, pseudonymised data that the controller does not hold the 'reidentification key' to. Controllers still need to safeguard such data, but have more relaxed obligations regarding the rights of the concerned individuals.
4. **Qualitative data** are difficult to anonymise because there may be indicators such as the combination of a location and an employment type which could make it easier to identify an individual or small cohort of individuals.
5. **Data analytics** describes a collection of data processing methods which use large amounts of data (*big data*) to derive models and predictions about future behaviours or activity. Data analytics introduce some risk of re-identification:
    - 5.1. **Cross-referencing or Cross-correlation:** when data are aggregated or correlated with other data, then the likelihood of being able to identify an individual, especially an outlier increases.
    - 5.2. **Co-morbidities:** for clinical data, where multiple conditions may present for an individual, this also increases the likelihood of being able to identify that individual.
6. **Statistical Disclosure Control** refers to methods used to reduce the risk of re-identification. They are encouraged when sharing or publishing data, and when publishing research outcomes (Willenborg & de Waal, 2001; Griffiths et al., 2019)

Our overall **Recommendations on Anonymity include:**

Check with your institution, data protection officer or authority, and institutional review board what local definitions of the terms are (e.g., *anonymous, pseudonymised, de-identified etc.*)

1. Check what the local (national) expectations are: *a data subject will usually expect their data to be processed in compliance with local instruments*

[7] Sometimes referred to as de-personalised

2. Check with the controller or data user what they claim the status of the data to be (*anonymous, de-identified, pseudonymised, etc.*). Nonetheless, as data identifiability can shift from jurisdiction to jurisdiction, and relative to the factual circumstances of its use, it is prudent not to rely on any representations made by third parties regarding the identifiability of their data.
3. Carry out a re-identification risk assessment before
    - 3.1. Combining one or more datasets
    - 3.2. Sharing or publishing data, or publishing research findings quoting examples of the data
4. Carry out an impact assessment[8] in regard to the impact on the data subject (the individual identified) before disclosure or publication, and introduce additional measures (Statistical Disclosure Control) to mitigate the risk.

9.4.5 Consent

Consent is the act by which a participant, patient or data subject indicates that they permit something to happen to them, or to their data, which would otherwise not be able to happen. It covers a number of different specific contexts:

1. **Clinical**: a patient agrees to undergoing a procedure, including taking part in a trial;
2. **Data Protection**: a data subject agrees to personal data being processed for specific purposes;
3. **Research**: a participant agrees to take part in a research study or experiment.

In both cases, the informed consent sheets for clinical or research purposes would explicitly set out how data protection will be handled, as well as samples or biobanking, rights to self-images and others

Giving consent should be informed (e.g. the individual knows what's going to happen and why), freely given (there is no coercion or similar motivation), given by somebody with capacity, unambiguous and auditable (the consent is recorded somewhere) (See also Parra-Calderón, 2018).

Ideally, consent should be sought for collecting, processing, sharing and publishing data. However, there are other legal bases for processing personal data. Some specific examples from the European General Data Protection Regulation (GDPR, 2016) are described below. Our recommendation would therefore be as follows:

1. Where possible, use data where the data subject has provided a valid consent that includes or is compatible with intended use of the data and complies with the requirements on consent in the specific country or region.

[8] What would be the impact to the data subject if they were identified from the data you hold.

Where these are not possible, there are other reasons why data may be used (see Hallinan 2020, Ó Cathaoir et al., 2020). For example, there may be a different legal basis for using personal data.

2. If using personal data, check whether there may be another basis for using the data.

In Europe, for instance, the GDPR provides other legal bases for processing personal data, we suggest:

1. **Vital Interests** (Art. 6(1)(d), and Art. 9(2)(c)): it may not be practical, feasible or possible to contact the data subject. However, to protect the *vital interests* of other natural persons the data needs to be interrogated and used.

In addition, there are other provisions for both personal data:

2. **Public Task** (Art. 6(1)(e))

and special category data:

3. **Public Interest** (Art. 9(2)(g))
4. **Preventive …Medicine** (Art. 9(2)(h))
5. **Public Health** (Art. 9(2)(i))
6. **Public Interest, Scientific or Historical Research Purposes or Statistical Purposes** (Art. 9(2)(j))

There is adequate provision, therefore, in the current regulation and its derivatives. In other jurisdictions, there may be other provisions which could be used. Their potential applicability in a specific case should be carefully examined.

9.4.6 The 5 Safes Model: Safe People, Safe Projects, Safe Settings, Safe Outputs, Safe Data

The 5 Safes Model was developed by staff working at the Office for National Statistics (UK) to be an easy to implement sensitive data management framework (Ritchie, 2008). It has subsequently been adopted by numerous Research Data Centres around the World.

The ambition of the 5 Safes Model is to achieve the 'Safe Use' of research data by accounting for five potential areas of risk to data subject confidentiality.

The 5 Safes are:
Safe People - Who is going to be accessing the data?

1. Safe People should have the right motivations for accessing research data.
2. Safe People should also have sufficient experience to work with the data safely.
3. Researchers may need to undergo specific training before using sensitive or confidential research data to become Safe People

Safe Projects - What is the purpose of accessing the data

1. Safe Projects are those that have a valid research purpose with a defined ‘public benefit’.
2. It must not be possible to realise this benefit without access to the data.

Safe Settings - Where will the data be accessed?

1. Access controls should be proportionate to the level of risk contained with the data.
2. Sensitive or confidential data should only be accessed via a suitable Safe Setting.
3. Safe Settings should have safeguards in place to minimise the risk that unauthorised people could access the data.

Safe Data - What does the data contain?

1. Safe Data will present minimal risk possible to the confidentiality of the data subjects..
2. The minimisation of risk could be achieved by removing direct identifiers, aggregating values, banding variables, or other statistical techniques that make re-identification more difficult. However, the loss of detail may limit the usefulness of the dataset.
3. Sensitive or confidential data should not be considered to be safe because of the residual risk to data subject confidentiality; however it is often the most useful for research.

Safe Outputs – What will be produced from the data?

1. Research that is generated from data may form derived outputs; these could include statistics, graphs/charts, or reports.
2. Outputs generated from the use of sensitive or confidential data should only be released if they report statistical findings and cannot be used to reveal the identity of a data subject nor enable the association of confidential information to a data subject.
3. Statistical Disclosure Control (SDC) is often used to minimise the risk of releasing confidential information.
4. Researchers and/or the institution managing the use of the data should check outputs (apply SDC) before publication to ensure they do not present undue risk. The intended outputs should have formed part of any application for ethical approval.

10. Additional Working Documents

Additional working documents about data sharing in clinical medicine	
Description of the resource	Link of the resource
Database of privately and publicly funded clinical studies conducted around the world	https://clinicaltrials.gov/
The Trans-NIH BioMedical Informatics Coordinating Committee (BMIC)	https://www.nlm.nih.gov/NIHbmic/index.html
Open-Access Data and Computational Resources to Address COVID-19	https://datascience.nih.gov/covid-19-open-access-resources
Research on “Sharing and reuse of individual participant data from clinical trials: principles and recommendations”	https://bmjopen.bmj.com/content/7/12/e018647
CDISC Interim User Guide for COVID-19	https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19
International COVID-19 Clinical Trials Map (based on the WHO Clinical Trials Search Portal)	https://covid-19.heigit.org/clinical_trials.html
ISO/TS 17975:2015:	https://www.iso.org/standard/61186.html
Official CDC guidelines for the new COVID-19 ICD-10-CM code	https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf
Additional information about clinical trials	https://docs.google.com/document/d/1LPj6AMS63f9rQziHGzvgD-5CEfrV5Tv_s8n-S1Fe6lI/edit
Additional information about clinical aspects	https://docs.google.com/document/d/1xNKVfzAxDFyqonop0SS-zxb-dOqKIyQkUIwlvsMABLU/edit

Additional working documents about community participation and data sharing	
Description of the resource	Link of the resource
Additional information about community participation	https://www.rd-alliance.org/groups/rda-covid19-community-participation

RDA-COVID-19 Community participation WG drafting	https://docs.google.com/document/d/1FEe2LIFR-D_yGR8Ow3LTrdYWCWo0ppC5YJrMDWFTGio/edit#heading=h.e5qs6lahfe5n
RDA COVID-19 WG Guidelines for Data Sharing	https://docs.google.com/document/d/1BqHrWfv__Jzr2YbuNaxIkW--4P9mMO1hwicmLqGMlmQ/edit?ts=5e95a561
	https://drive.google.com/drive/folders/1u6FpmTP5X6Vw3Hjtl1I__IT-lgmbwv7w

Additional working documents about data sharing in epidemiology	
Description of the resource	Link of the resource
Additional information about data sharing in epidemiology	https://www.rd-alliance.org/groups/rda-covid19-epidemiology

Additional working documents about data sharing in omics practices	
Description of the resource	Link of the resource
Additional information about data sharing in omics practices	https://www.rd-alliance.org/groups/rda-covid19-omics

Additional working documents about data sharing in social sciences	
Description of the resource	Link of the resource
Complete guidelines	https://docs.google.com/document/d/1TmMdq-93cAquy0DbV90XO0GfLu5JfI6mpgkt6KewjK8/edit
Additional information about data sharing in social sciences	https://www.rd-alliance.org/groups/rda-covid19-social-sciences
Best practices for software and trustworthy analysis	https://docs.google.com/document/d/14Cd1cOS8Cv8HhLEkVyPv2UfunvHdLsUaNk7U7AdDJk8/edit?usp=sharing

Additional working documents about research software and data sharing	
Description of the resource	Link of the resource
Additional information about research software and data sharing	https://www.rd-alliance.org/groups/rda-covid19-software

Additional working documents about legal and ethical compliance	
Description of the resource	Link of the resource
Additional information about legal and ethical compliance	https://www.rd-alliance.org/groups/rda-covid19-legal-ethical

11. References

AAS. "Statement on COVID-19: Ethics, Governance and Community Engagement in Times of Crises." African Academy of Sciences, Biospecimens and Data Governance Committee, 2020. https://www.aasciences.africa/sites/default/files/2020-04/Covid-19%20Ethics%2C%20Governance%20%26%20Community%20Engagement%20in%20times%20of%20crisis_0.pdf.

Access Now. "Recommendations on Privacy and Data Protection in the Fight against COVID-19," March 2020. https://www.accessnow.org/cms/assets/uploads/2020/03/Access-Now-recommendations-on-Covid-and-data-protection-and-privacy.pdf.

Addshore, Daniel Mietchen, Egon Willighagen, and Yayamamo. *SARS-CoV-2-Queries*, 2020. https://egonw.github.io/SARS-CoV-2-Queries/.

Aguilar-Gallegos, Norman, Leticia Elizabeth Romero-García, Enrique Genaro Martínez-González, Edgar Iván García-Sánchez, and Jorge Aguilar-Ávila. "Dataset on Dynamics of Coronavirus on Twitter." *Data in Brief*, May 2020, 105684. https://doi.org/10.1016/j.dib.2020.105684.

AIRR Community. "Adaptive Immune Receptor Repertoire - Data Commons API V1 - AIRR Standards 1.3.0 Documentation." Adaptive Immune Receptor Repertoire (AIRR) - Common Repository Working Group (CRWG) - AIRR Data Commons API V1 — AIRR Standards 1.3.0 documentation, 2020. https://docs.airr-community.org/en/latest/api/adc_api.html.

———. "MiAIRR-to-NCBI Implementation." AIRR Standards 1.3.0 documentation, 2020. https://docs.airr-community.org/en/latest/miairr/miairr_ncbi_overview.html.

Alibaba Cloud. "Free Computational and AI Platforms to Help Research, Analyze and Combat COVID-19 ('Program')." Elastic HPC Solution for Life Sciences on COVID-19 Research - Alibaba Cloud, 2020. https://www.alibabacloud.com/solutions/lifesciences-ehpc.

ALLEA. "The European Code of Conduct for Research Integrity," 2017. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf.

Artic Network. "Artic Network: HCoV-2019 (NCoV-2019/SARS-CoV-2)." Artic Network, 2020. https://artic.network/ncov-2019.

Article 29 Working Party on Data Protection. "Article 29 Data Protection Working Party Comments in Response to W3C's Public Consultation on the W3C Last Call Working Draft, 14 July 2015, Tracking Compliance and Scope," October 1, 2015. https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20151001__letter_of_the_art_29_wp_w3c_compliance.pdf.

———. "Opinion 4/2007 on the Concept of Personal Data," June 20, 2007. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf.

———. "Opinion 05/2014 on Anonymisation Techniques," April 10, 2015. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

Askitas, Nikolaos. "A Data Tax for a Digital Economy." World of Labor IZA, October 22, 2018. https://wol.iza.org/opinions/a-data-tax-for-a-digital-economy.

Athar, Awais, Anja Füllgrabe, Nancy George, Haider Iqbal, Laura Huerta, Ahmed Ali, Catherine Snow, et al. "ArrayExpress Update – from Bulk to Single-Cell Expression Data." *Nucleic Acids Research* 47, no. D1 (January 8, 2019): D711–15. https://doi.org/10.1093/nar/gky964.

Barrett, Tanya, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, et al. "NCBI GEO: Archive for Functional Genomics Data Sets—Update." *Nucleic Acids Research* 41, no. D1 (January 1, 2013): D991–95. https://doi.org/10.1093/nar/gks1193.

Battegay, Manuel, Richard Kuehl, Sarah Tschudin-Sutter, Hans H. Hirsch, Andreas F. Widmer, and Richard A. Neher. "2019-Novel Coronavirus (2019-NCoV): Estimating the Case Fatality Rate – a Word of Caution." *Swiss Medical Weekly* 150, no. 0506 (February 7, 2020). https://doi.org/10.4414/smw.2020.20203.

BBMRI-NL. "Integrative Omics Data Set | BBMRI." Bio Banking Netherlands, 2020. https://bbmri.nl/services/samples-images-data/integrative-omics-data-set.

Beck, T, T Shorter, and AJ Brookes. "GWAS Central," 2020. https://www.gwascentral.org/.

Berman, Helen, Kim Henrick, and Haruki Nakamura. "Announcing the Worldwide Protein Data Bank." *Nature Structural & Molecular Biology* 10, no. 12 (December 2003): 980–980. https://doi.org/10.1038/nsb1203-980.

Berman, Helen M., John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. "The Protein Data Bank." *Nucleic Acids Research* 28, no. 1 (January 1, 2000): 235–42. https://doi.org/10.1093/nar/28.1.235.

Bernstein, Frances C., Thomas F. Koetzle, Graheme J.B. Williams, Edgar F. Meyer, Michael D. Brice, John R. Rodgers, Olga Kennard, Takehiko Shimanouchi, and Mitsuo Tasumi. "The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures." *Journal of Molecular Biology* 112, no. 3 (May 1977): 535–42. https://doi.org/10.1016/S0022-2836(77)80200-3.

Bernstein, H. J., J. C. Bollinger, I. D. Brown, S. Gražulis, J. R. Hester, B. McMahon, N. Spadaccini, J. D. Westbrook, and S. P. Westrip. "Specification of the Crystallographic Information File Format, Version 2.0." *Journal of Applied Crystallography* 49, no. 1 (February 1, 2016): 277–84. https://doi.org/10.1107/S1600576715021871.

Berry, Isha, Jean-Paul R. Soucy, Ashleigh Tuite, and David Fisman. "Open Access Epidemiologic Data and an Interactive Dashboard to Monitor the COVID-19 Outbreak in Canada." *Canadian Medical Association Journal* 192, no. 15 (April 14, 2020): E420–E420. https://doi.org/10.1503/cmaj.75262.

Bhattacharya, Sanchita, Patrick Dunn, Cristel G. Thomas, Barry Smith, Henry Schaefer, Jieming Chen, Zicheng Hu, et al. "ImmPort, toward Repurposing of Open Access Immunological Assay Data for Translational and Clinical Research." *Scientific Data* 5 (27 2018): 180015. https://doi.org/10.1038/sdata.2018.15.

BioExcel, and The Molecular Sciences Software Institute (MolSSI). "COVID-19 Molecular Structure and Therapeutics Hub." COVID-19 Molecular Structure and Therapeutics Hub, 2020. https://covid.bioexcel.eu/.

Blaxter, Mark, Antoine Danchin, Babis Savakis, Kaoru Fukami-Kobayashi, Ken Kurokawa, Sumio Sugano, Richard J. Roberts, Steven L. Salzberg, and Chung-I. Wu. "Reminder to Deposit DNA Sequences." *Science* 352, no. 6287 (May 13, 2016): 780–780. https://doi.org/10.1126/science.aaf7672.

Broad Institute. "Genotype-Tissue Expression (GTEx) Portal." The Broad Institute of MIT and

Harvard, 2020. https://www.gtexportal.org/home/.

Brown, D. A., M. B. Chadwick, R. Capote, A. C. Kahler, A. Trkov, M. W. Herman, A. A. Sonzogni, et al. "ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-Project Cross Sections, New Standards and Thermal Scattering Data." *Nuclear Data Sheets*, Special Issue on Nuclear Reaction Data, 148 (February 1, 2018): 1–142. https://doi.org/10.1016/j.nds.2018.02.001.

Cabellos, O., F. Alvarez-Velarde, M. Angelone, C. J. Diez, J. Dyrda, L. Fiorito, U. Fischer, et al. "Benchmarking and Validation Activities within JEFF Project." *EPJ Web of Conferences* 146 (2017): 06004. https://doi.org/10.1051/epjconf/201714606004.

Carlson, A. D., V. G. Pronyaev, R. Capote, G. M. Hale, Z. -P. Chen, I. Duran, F. -J. Hambsch, et al. "Evaluation of the Neutron Data Standards." *Nuclear Data Sheets*, Special Issue on Nuclear Reaction Data, 148 (February 1, 2018): 143–88. https://doi.org/10.1016/j.nds.2018.02.002.

CCSDS. "Audit and Certification of Trustworthy Digital Repositories." Consultative Committee for Space Data Systems, 2011.

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

———. "Public Health and Promoting Interoperability Programs (Formerly, Known as Electronic Health Records Meaningful Use)," March 24, 2020. https://www.cdc.gov/ehrmeaningfuluse/introduction.html.

———. "Weekly Provisional Death Counts by Select Demographic and Geographic Characteristics." Dataset. Center for Disease Control, USA, 2020. https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm.

CDISC. "CDISC Standards in the Clinical Research Process." Text. CDISC, 2020. https://www.cdisc.org/standards.

———. "Interim User Guide for COVID-19." Text. CDISC, 2020. https://www.cdisc.org/standards/therapeutic-areas/covid-19.

CESSDA. "Adapt Your Data Management Plan." CESSDA, 2019. https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf.

———. "Adapt Your DMP (Editable Form)." CESSDA, 2018. https://www.cessda.eu/content/download/4304/48666/file/TTT_DO_DMPExpertGuideEditVersion_v1.2.docx.

Chan, Andrew T., David A. Drew, Long H. Nguyen, Amit D. Joshi, Wenjie Ma, Chuan-Guo Guo, Chun-Han Lo, et al. "The COronavirus Pandemic Epidemiology (COPE) Consortium: A Call to Action." *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, May 5, 2020. https://doi.org/10.1158/1055-9965.EPI-20-0606.

Chang, Christopher C., Carson C. Chow, Laurent Cam Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4 (2015): 7. https://doi.org/10.1186/s13742-015-0047-8.

Christley, Scott, Walter Scarborough, Eddie Salinas, William H. Rounds, Inimary T. Toby, John M. Fonner, Mikhail K. Levin, et al. "VDJServer: A Cloud-Based Analysis Portal and Data Commons for Immune Repertoire Sequences and Rearrangements." *Frontiers in*

Immunology 9 (2018): 976. https://doi.org/10.3389/fimmu.2018.00976.

Clinical Data Interchange Standards Consortium (CDISC). "CDISC Interim User Guide for COVID-19 - CDISC Interim User Guide for COVID-19 - Wiki." Clinical Data Interchange Standards Consortium, Inc., April 21, 2020. https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19.

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

Cock, Peter J. A., Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. "The Sanger FASTQ File Format for Sequences with Quality Scores, and the Solexa/Illumina FASTQ Variants." *Nucleic Acids Research* 38, no. 6 (April 1, 2010): 1767–71. https://doi.org/10.1093/nar/gkp1137.

Coffey, Barbara. "LibGuides: Finance: Info on Company Ids and Linking Data Sources." Accessed May 14, 2020. https://libguides.princeton.edu/c.php?g=939414&p=6776005.

CoreTrustSeal. "CoreTrustSeal." CoreTrustSeal. Accessed May 10, 2020. https://www.coretrustseal.org/.

CoreTrustSeal Standards and Certification Board. "CoreTrustSeal Trustworthy Data Repositories Requirements: Extended Guidance 2020–2022," November 20, 2019. https://doi.org/10.5281/zenodo.3632532.

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

Corrie, Brian D., Nishanth Marthandan, Bojan Zimonja, Jerome Jaglale, Yang Zhou, Emily Barr, Nicole Knoetze, et al. "IReceptor: A Platform for Querying and Analyzing Antibody/B-Cell and T-Cell Receptor Repertoire Data across Federated Repositories." *Immunological Reviews* 284, no. 1 (2018): 24–41. https://doi.org/10.1111/imr.12666.

COS. "Coronavirus Outbreak Research Collection." Center for Open Science, 2020. https://osf.io/collections/coronavirus/discover.

Council of Europe. Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, Pub. L. No. Treaty No.164 (1999). https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168007cf98.

———. "European Convention on Human Rights," June 1, 2010. https://www.echr.coe.int/Documents/Convention_ENG.pdf.

———. "The Impact of the COVID-19 Pandemic on Human Rights and the Rule of Law," April 2020. https://www.coe.int/en/web/human-rights-rule-of-law/covid19.

Council of Europe Bioethics. "COVID-19," 2020. https://www.coe.int/en/web/bioethics/covid-19.

———. "DH-BIO Statement on Human Rights Considerations Relevant to the COVID-19 Pandemic," April 14, 2020. https://rm.coe.int/inf-2020-2-statement-covid19-e/16809e2785.

COVID-19 Biohackathon April 5-11 Participants, University of Manchester, and HITS gGmbH. "A COVID-19-Specific Instance for EOSC-Life's WorkflowHub." The WorkflowHub, 2020. https://covid19.workflowhub.eu/.

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

Creative Commons. "CC BY 4.0 - Attribution 4.0 International," 2020. https://creativecommons.org/licenses/by/4.0/.

———. "CC0 1.0 - Universal," 2020. https://creativecommons.org/publicdomain/zero/1.0/.

———. "Creative Commons — Attribution 4.0 International — CC BY 4.0." Accessed May 10, 2020. https://creativecommons.org/licenses/by/4.0/.

———. "Creative Commons — CC0 1.0 Universal." Accessed May 10, 2020. https://creativecommons.org/publicdomain/zero/1.0/.

Cross Section Evaluation Working Group. "CSEWG," 2020. https://www.nndc.bnl.gov/csewg/.

Davis, Larry. *Corona Data Scraper*. HTML. 2020. Reprint, COVID Atlas, 2020. https://github.com/covidatlas/coronadatascraper.

DCC. "DMP Online." Digital Curation Centre. Accessed April 29, 2020. https://dmponline.dcc.ac.uk/.

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

———. "DDBJ - BioSample." DDBJ BioSample - Home, February 19, 2018. /biosample/index-e.html.

———. "DNA Data Bank of Japan (DDBJ)," September 30, 2019. https://www.ddbj.nig.ac.jp/.

DDI Alliance. "Data Documentation Initiative," 2020. https://ddialliance.org/.

DeBord, D. Gayle, Tania Carreon, and Thomas Lentz. "Use of the 'Exposome' in the Practice of Epidemiology: A Primer on -Omic Technologies." *Am J Epidemiology* 184, no. 4 (2016): 312–14. https://doi.org/10.1093/aje/kwv325.

Deutsch, Eric W. "The PeptideAtlas Project." Edited by Simon J. Hubbard and Andrew R. Jones. *Proteome Bioinformatics*, 2010, 285–96. https://doi.org/10.1007/978-1-60761-444-9_19.

Deutsch, Eric W., Nuno Bandeira, Vagisha Sharma, Yasset Perez-Riverol, Jeremy J. Carver, Deepti J. Kundu, David García-Seisdedos, et al. "The ProteomeXchange Consortium in 2020: Enabling 'big Data' Approaches in Proteomics." *Nucleic Acids Research* 48, no. D1 (08 2020): D1145–52. https://doi.org/10.1093/nar/gkz984.

Djalante, Riyanti, Rajib Shaw, and Andrew DeWit. "Building Resilience against Biological Hazards and Pandemics: COVID-19 and Its Implications for the Sendai Framework." *Progress in Disaster Science* 6 (April 1, 2020): 100080. https://doi.org/10.1016/j.pdisas.2020.100080.

DNAstack. "COVID-19 Beacon." COVID-19 Beacon, 2020. https://covid-19.dnastack.com/_/discovery?position=3840&referenceBases=A&alternateBases=G.

Dong, Ensheng, Hongru Du, and Lauren Gardner. "An Interactive Web-Based Dashboard to Track COVID-19 in Real Time." *The Lancet. Infectious Diseases* 20, no. 5 (May 2020): 533–34. https://doi.org/10.1016/S1473-3099(20)30120-1.

Drew, David A., Long H. Nguyen, Claire J. Steves, Cristina Menni, Maxim Freydin, Thomas Varsavsky, Carole H. Sudre, et al. "Rapid Implementation of Mobile Technology for Real-Time Epidemiology of COVID-19." *Science (New York, N.Y.)*, May 5, 2020. https://doi.org/10.1126/science.abc0473.

DSI, DG SANTE, CEF eHealth. "EHDSI INTEROPERABILITY SPECIFICATIONS, Requirements and Frameworks (Normative Artefacts) - EHealth DSI Operations - CEF Digital," March 24, 2020.

https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=35210463.

DTL. "Personal Health Train." Dutch Techcentre for Life Sciences, 2018. https://www.dtls.nl/fair-data/personal-health-train/.

ECDC. "Geographic Distribution of COVID-19 Cases Worldwide." Dataset. European Centre for Disease Prevention and Control, April 15, 2020. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.

———. "The European Surveillance System (TESSy)." European Centre for Disease Prevention and Control, 2019. https://www.ecdc.europa.eu/en/publications-data/european-surveillance-system-tessy.

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

EMBL-EBI. "ArrayExpress." ArrayExpress < EMBL-EBI, 2020. https://www.ebi.ac.uk/arrayexpress/.

———. "European Nucleotide Archive (ENA)." European Nucleotide Archive < EMBL-EBI, 2008. https://www.ebi.ac.uk/ena.

———. "Expression Atlas: Gene Expression across Species and Biological Conditions." European Molecular Biology Laboratory European Bioinformatics Institute, 2020. https://www.ebi.ac.uk/gxa/home.

———. "Pathogens: Surveillance, Identification Investigation." European Molecular Biology Laboratory - European Bioinformatics Institute, 2020. https://www.ebi.ac.uk/ena/pathogens/covid-19.

ENA-Docs. "ENA Documentation," 2020. https://ena-docs.readthedocs.io/en/latest/.

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

EU. "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, Ev_20190611_co922_en.Pdf." EU eHealth Network, 2019. https://ec.europa.eu/health/sites/health/files/ehealth/docs/ev_20190611_co922_en.pdf.

European Commission. COMMISSION RECOMMENDATION (EU) 2020/518 of 8 April 2020 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 anonymised mobility data (2020). https://ec.europa.eu/info/sites/info/files/recommendation_on_apps_for_contact_tracing_4.pdf.

———. "European Reference Networks." Text. Public Health - European Commission, 2016. https://ec.europa.eu/health/ern/networks_en.

———. "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, April 18, 2020. https://www.rd-alliance.org/system/files/documents/H2020_Guidelines_COVID19_EC.pdf.

———. "Pseudonymisation Tool." EUPID - European Platform on Rare Disease Registration,

2020. https://eu-rd-platform.jrc.ec.europa.eu.

———. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Data Protection Board. "Statement on the Processing of Personal Data in the Context of the COVID-19 Outbreak," March 19, 2020. https://edpb.europa.eu/our-work-tools/our-documents/other/statement-processing-personal-data-context-covid-19-outbreak_en.

European Group on Ethics in Science and New Technologies. "Statement on European Solidarity and the Protection of Fundamental Rights in the COVID-19 Pandemic," April 2, 2020. https://ec.europa.eu/info/sites/info/files/research_and_innovation/ege/ec_rtd_ege-statement-covid-19.pdf.

———. "Statement on European Solidarity and the Protection of Fundamental Rights in the COVID-19 Pandemic," April 2, 2020.

European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI). "EMBL-EBI COVID-19 Data Portal," 2020. https://www.covid19dataportal.org/.

European Nucleotide Archive. "ENA Virus Pathogen Reporting Standard Checklist." EMBL-EBI, 2020. https://www.ebi.ac.uk/ena/data/view/ERC000033.

———. "ENA Virus Pathogen Reporting Standard Checklist," April 21, 2020. https://www.ebi.ac.uk/ena/data/view/ERC000033.

———. "SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) Submissions — Ena-Browser-Docs Latest Documentation," 2020. https://ena-browser-docs.readthedocs.io/en/latest/help_and_guides/sars-cov-2-submissions.html.

FAIR4Health. "FAIR4Health at RDA Germany Conference 2020 - Resources." FAIR4Health Consortium, 2020. https://www.fair4health.eu/en/resources.

FAIRsharing. "AIRR Rearrangement File Format." FAIRsharing, 2020. https://fairsharing.org/bsg-s001474.

———. "Binary Alignment Map Format." FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.HZA1EC.

———. "Browser Extensible Data Format." FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.MWMBPQ.

———. "DNA Data Bank of Japan." FAIRsharing, 2020. https://doi.org/10.25504/FAIRsharing.k337f0.

———. "EMBL-EBI - BioSamples," May 11, 2020. https://doi.org/10.25504/FAIRsharing.ewjdq6.

———. "European Genome-Phenome Archive." FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.MYA1FF.

———. "European Nucleotide Archive." FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.DJ8NT8.

———. "FAIRsharing." Accessed May 14, 2020. https://fairsharing.org/.

———. "FAIRsharing - Adaptive Immune Receptor Repertoire Resources," 2020. https://fairsharing.org/search/?q=AIRR.

———. "FAIRsharing - Genomics Databases," 2020. https://fairsharing.org/search/?q=genomics&content=biodbcore.

———. "FAIRsharing - Genomics Standards," 2020. https://fairsharing.org/search/?q=genomics&content=standards.

———. “FAIRsharing - Transcriptomics Databases,” 2020. https://fairsharing.org/search/?q=transcriptomics&content=biodbcore.

———. “FAIRsharing - Transcriptomics Standards,” 2020. https://fairsharing.org/search/?q=transcriptomics&content=standards.

———. “FAIRsharing Collection: COVID-19 Resources.” FAIRsharing Collection: COVID-19 Resources, 2020. https://fairsharing.org/collection/COVID19Resources.

———. “FAIRsharing Record for: ArrayExpress.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.6K0KWD.

———. “FAIRsharing Record for: Gene Expression Omnibus.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.5HC8VT.

———. “FASTA Sequence Format.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.RZ4VFG.

———. “FASTQ Sequence and Sequence Quality Format.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.R2TS5T.

———. “Flow Cytometry Data File Standard.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.QRR33Y.

———. “Gating-ML.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.QPYP5G.

———. “Gene Transfer Format.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.SGGB1N.

———. “Generic Feature Format Version 3.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.DNK0F6.

———. “Genome-Wide Association Studies Catalog.” FAIRsharing, 2018. https://doi.org/10.25504/FAIRSHARING.BLUMRX.

———. “Genomic Expression Archive.” FAIRsharing, 2018. https://doi.org/10.25504/FAIRSHARING.HESBCY.

———. “GWAS Central.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.VKR57K.

———. “Japanese Genotype-Phenotype Archive.” FAIRsharing, 2018. https://doi.org/10.25504/FAIRSHARING.PWGF4P.

———. “Minimal Information about a High Throughput SEQuencing Experiment.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.A55Z32.

———. “Minimal Information about Adaptive Immune Receptor Repertoire.” FAIRsharing, 2018. https://doi.org/10.25504/FAIRSHARING.31HEC1.

———. “Minimum Information about Flow Cytometry.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.KCNJJ2.

———. “NCBI - BioSamples,” January 8, 2019. https://doi.org/10.25504/FAIRsharing.qr6pqk.

———. “NCBI GenBank.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.9KAHY4.

———. “NCBI Sequence Read Archive.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.G7T2HV.

———. “NCBI Viral Genomes.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.QT5KY7.

———. “Sequence Alignment Map.” FAIRsharing, 2015. https://doi.org/10.25504/FAIRSHARING.K97XZH.

———. “Variant Call Format.” FAIRsharing, 2015.

https://doi.org/10.25504/FAIRSHARING.CFZZ0H.

———. "VDJServer." FAIRsharing, 2018. https://doi.org/10.25504/FAIRSHARING.NZDQ0F.

Farrah, Terry, Eric W. Deutsch, Richard Kreisberg, Zhi Sun, David S. Campbell, Luis Mendoza, Ulrike Kusebauch, et al. "PASSEL: The PeptideAtlas SRMexperiment Library." *Proteomics* 12, no. 8 (April 2012): 1170–75. https://doi.org/10.1002/pmic.201100515.

FDA. "Sentinel Common Data Model | Sentinel Initiative." FDA Sentinel Initiative, 2019. https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model.

Felsenstein, Joe. "The Newick Tree Format." The Newick tree format, 1986. http://evolution.genetics.washington.edu/phylip/newicktree.html.

Finnie, Thomas, Andy South, and Ana Bento. "EpiJSON: A Unified Data-Format for Epidemiology." *Epidemics* 15, no. June, 2016 (2016): 20–26. https://doi.org/10.1016/j.epidem.2015.12.002.

Fitzgerald, P. M. D., J. D. Westbrook, P. E. Bourne, B. McMahon, K. D. Watenpaugh, and H. M. Berman. "Macromolecular Dictionary (MmCIF)." In *International Tables for Crystallography*, edited by S. R. Hall and B. McMahon, 1st ed., G:295–443. International Tables for Crystallography. Chester, England: International Union of Crystallography, 2006. https://doi.org/10.1107/97809553602060000745.

FitzHenry, F., F.S. Resnic, S.L. Robbins, J. Denton, L. Nookala, D. Meeker, L. Ohno-Machado, and M.E. Matheny. "Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership." *Applied Clinical Informatics* 6, no. 3 (2015): 536–47. https://doi.org/10.4338/ACI-2014-12-CR-0121.

FORCE11. "Guiding Principles for Findable, Accessible, Interoperable and Re-Usable Data Publishing Version B1.0," 2017. https://www.force11.org/fairprinciples.

Freunde von GISAID e.V. ("GISAID"). "GISAID: Genomic Epidemiology of HCoV-19." GISAID - Next hCoV-19 App, 2020. https://www.gisaid.org/epiflu-applications/next-hcov-19-app/.

Fritz, Markus Hsi-Yang, Rasko Leinonen, Guy Cochrane, and Ewan Birney. "Efficient Storage of High Throughput DNA Sequencing Data Using Reference-Based Compression." *Genome Research* 21, no. 5 (May 1, 2011): 734–40. https://doi.org/10.1101/gr.114819.110.

GA4GH. "Enabling Responsible Genomic Data Sharing for the Benefit of Human Health." Global Alliance for Genomics and Health, 2020. https://www.ga4gh.org/.

———. "GA4GH: Data Security Toolkit." Global Alliance for Genomics and Health, 2020. https://www.ga4gh.org/genomic-data-toolkit/data-security-toolkit/.

———. "GA4GH: Genomic Data Toolkit." Global Alliance for Genomics and Health, 2020. https://www.ga4gh.org/genomic-data-toolkit/.

———. "GA4GH: Regulatory & Ethics Toolkit." Global Alliance for Genomics and Health, 2020. https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/.

Galaxy Project. "Best Practices for the Analysis of SARS-CoV-2 Data: Genomics, Evolution, and Cheminformatics." COVID-19 analysis on usegalaxy, 2020. https://covid19.galaxyproject.org/.

GenBank. "SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) Sequences." U.S. Center for Disease Control, 2020. https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/.

German Data Forum (RatSWD). "Remote Access to Data from Official Statistics Agencies and Social Security Agencies." *RatSWD Output Paper Series*, 2020. https://doi.org/10.17620/02671.48.

Ghani, A. C., C. A. Donnelly, D. R. Cox, J. T. Griffin, C. Fraser, T. H. Lam, L. M. Ho, et al. "Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease." *American Journal of Epidemiology* 162, no. 5 (2005): 479–86. https://doi.org/10.1093/aje/kwi230.

Global Alliance for Genomics & Health. "Framework for Responsible Sharing of Genomic and Health-Related Data," n.d. 2014-12-09. https://www.ga4gh.org/genomic-data-toolkit/regulatory-ethics-toolkit/framework-for-responsible-sharing-of-genomic-and-health-related-data/.

Global Alliance for Genomics & Health, and 1000 Genomes Project. *Samtools/Hts-Specs*. TeX. 2012. Reprint, samtools, 2020. https://github.com/samtools/hts-specs.

Global Alliance for Genomics and Health (GA4GH). "Global Alliance for Genomics and Health Consent Policy." Global Alliance for Genomics and Health (GA4GH), September 2019. https://www.ga4gh.org/wp-content/uploads/GA4GH-Final-Revised-Consent-Policy_16Sept2019.pdf.

Global Health Drug Discovery Institute (GHDDI). "Targeting COVID-19: GHDDI Info Sharing Portal." Home - Targeting COVID-19 Portal, May 6, 2020. https://ghddi-ailab.github.io/Targeting2019-nCoV/.

Global Indigenous Data Alliance. "GIDA." GIDA Global Indigenous Data Alliance Promoting Indigenous Control of Indigenous Data, 2019. https://www.gida-global.org/.

GLOPID, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. "Principles of Data Sharing in Public Health Emergencies," 2018. https://www.glopid-r.org/wp-content/uploads/2018/06/glopid-r-principles-of-data-sharing-in-public-health-emergencies.pdf.

Goni, Ramon, Magnus Lundborg, Christoph Bernau, Ferdinand Jamitzky, Erwin Laure, Yolanda Becerra, Modesto Orozco, and Josep Lluís Gelpi. "Standards for Data Handling," 2013. https://www.bsc.es/sites/default/files/public/life_science/molecular_modeling/d7.3_-_white_paper_on_standards_for_data_handling.pdf.

Google. "Google Dataset Search," 2020. https://datasetsearch.research.google.com/.

Griffiths, Emily, Carlotta Greci, Yannis Kotrotsios, Simon Parker, James Scott, Richard Welpton, Arne Wolters, and Christine Woods. "Handbook on Statistical Disclosure Control for Outputs," 2019. https://doi.org/10.6084/m9.figshare.9958520.v1.

GTF2.2: A Gene Annotation Format. "GTF2.2: A Gene Annotation Format," 2003. https://mblab.wustl.edu/GTF22.html.

Guo, FB, and CT Zhang. "VGAS (Viral Genome Annotation System)," 2020. http://cefg.uestc.cn/vgas/.

Hall, S. R., F. H. Allen, and I. D. Brown. "The Crystallographic Information File (CIF): A New Standard Archive File for Crystallography." *Acta Crystallographica Section A* 47, no. 6 (November 1991): 655–685. https://doi.org/10.1107/S010876739101067X.

Hallinan, Dara. "Broad Consent under the GDPR: An Optimistic Perspective on a Bright Future." *Life Sciences, Society and Policy* 16, no. 1 (January 6, 2020). https://doi.org/10.1186/s40504-019-0096-3.

Han, Mira V., and Christian M. Zmasek. "PhyloXML: XML for Evolutionary Biology and

Comparative Genomics." *BMC Bioinformatics* 10, no. 1 (October 27, 2009): 356. https://doi.org/10.1186/1471-2105-10-356.

Hare, S.S., J.C.L. Rodrigues, J. Jacob, A. Edey, A. Devaraj, A. Johnstone, R. McStay, A. Nair, and G. Robinson. "A UK-Wide British Society of Thoracic Imaging COVID-19 Imaging Repository and Database: Design, Rationale and Implications for Education and Research." *Clinical Radiology* 75, no. 5 (May 2020): 326–28. https://doi.org/10.1016/j.crad.2020.03.005.

Hatcher, Eneida L., Sergey A. Zhdanov, Yiming Bao, Olga Blinkova, Eric P. Nawrocki, Yuri Ostapchuck, Alejandro A. Schäffer, and J. Rodney Brister. "Virus Variation Resource - Improved Response to Emergent Viral Outbreaks." *Nucleic Acids Research* 45, no. D1 (2017): D482–90. https://doi.org/10.1093/nar/gkw1065.

Haug, Kenneth, Keeva Cochrane, Venkata Chandrasekhar Nainala, Mark Williams, Jiakang Chang, Kalai Vanii Jayaseelan, and Claire O'Donovan. "MetaboLights: A Resource Evolving in Response to the Needs of Its Scientific Community." *Nucleic Acids Research* 48, no. D1 (January 8, 2020): D440–44. https://doi.org/10.1093/nar/gkz1019.

Hausman, Jessica, Shelley Stall, James Gallagher, and Mingfang Wu. "Software and Services Citation Guidelines and Examples." ESIP, February 20, 2019. https://esip.figshare.com/articles/Software_and_Services_Citation_Guidelines_and_Examples/7640426.

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

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

Heller, Stephen R., Alan McNaught, Igor Pletnev, Stephen Stein, and Dmitrii Tchekhovskoi. "InChI, the IUPAC International Chemical Identifier." *Journal of Cheminformatics* 7, no. 1 (May 30, 2015): 23. https://doi.org/10.1186/s13321-015-0068-4.

HL7. "HL7 Standards Product Brief - CDA® Release 2 | HL7 International," 2010. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=7.

HLA Covid-19 Consortium. "HLA COVID-19," 2020. http://hlacovid19.org/.

Hoffman, Andreas. "CApp: Chemical Data Formats." Hofmann Laboratory, 2020. http://www.structuralchemistry.org/pcsb/capp_cdf.php#sdf.

HUPO Proteomics Standards Initiative. "MzML 1.1.0 Specification | HUPO Proteomics Standards Initiative." mzML 1.1.0 Specification | HUPO Proteomics Standards Initiative, November 3, 2017. http://www.psidev.info/mzML.

———. "The Minimum Information About a Proteomics Experiment (MIAPE)," 2007. http://www.psidev.info/miape.

IHME. "COVID-19 Projections." Seattle, Washington, USA: Institute for Health Metrics and Evaluation, 2020. https://covid19.healthdata.org/projections.

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

Illumina. "FASTQ Files Explained." FASTQ files explained, March 10, 2020. https://support.illumina.com/bulletins/2016/04/fastq-files-explained.html.

INSEAD'. "INSEAD Research & Learning Hub." INSEAD, January 28, 2016. https://www.insead.edu/library/research/company-identifiers.

International Nucleotide Sequence Database Collaboration. "GenBank Overview." GenBank Overview, 2013. https://www.ncbi.nlm.nih.gov/genbank/.

International Nucleotide Sequence Database Collaboration (INSDC). "International Nucleotide Sequence Database Collaboration (INSDC)." International Nucleotide Sequence Database Collaboration | INSDC, 2020. http://www.insdc.org/.

International Organization for Standardization. "ISO/TS 17975:2015." ISO, 2015. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/11/61186.html.

International Union of Crystallography. "Catalogue of Metadata Resources for Crystallographic and Related Applications." (IUCr) metadata catalogue, 2020. https://www.iucr.org/resources/data/dddwg/metadata-catalogue.

International Union of Crystallography (IUCr). "Crystallographic Information Framework (CIF)." (IUCr) Crystallographic Information Framework, 1991. https://www.iucr.org/resources/cif.

Italian Civil Protection Department, Micaela Morettini, Agnese Sbrollini, Ilaria Marcantoni, and Laura Burattini. "COVID-19 in Italy: Dataset of the Italian Civil Protection Department." *Data in Brief* 30 (June 2020): 105526. https://doi.org/10.1016/j.dib.2020.105526.

Janes, Jeff, Megan E. Young, Emily Chen, Nicole H. Rogers, Sebastian Burgstaller-Muehlbacher, Laura D. Hughes, Melissa S. Love, et al. "The ReFRAME Library as a Comprehensive Drug Repurposing Library and Its Application to the Treatment of Cryptosporidiosis." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 42 (16 2018): 10750–55. https://doi.org/10.1073/pnas.1810137115.

JHU. "COVID19 Dataset." Dataset. 2020. Reprint, Johns Hopkins University, CSSEGISandData, April 12, 2020. https://github.com/CSSEGISandData/COVID-19.

Kabach, Ouadie, Abdelouahed Chetaine, and Abdelfettah Benchrif. "Processing of JEFF-3.3 and ENDF/B-VIII.0 and Testing with Critical Benchmark Experiments and TRIGA Mark II Research Reactor Using MCNPX." *Applied Radiation and Isotopes* 150 (August 1, 2019): 146–56. https://doi.org/10.1016/j.apradiso.2019.05.015.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and and David Haussler. "The Human Genome Browser at UCSC." *Genome Research* 12, no. 6 (June 1, 2002): 996–1006. https://doi.org/10.1101/gr.229102.

Kinjo, Akira R., Gert-Jan Bekker, Hirofumi Suzuki, Yuko Tsuchiya, Takeshi Kawabata, Yasuyo Ikegawa, and Haruki Nakamura. "Protein Data Bank Japan (PDBj): Updated User Interfaces, Resource Description Framework, Analysis Tools for Large Structures." *Nucleic Acids Research* 45, no. D1 (October 26, 2016): D282–88. https://doi.org/10.1093/nar/gkw962.

Knight, Gwenan, Nila Dharan, and Gregory Fox. "Bridging the Gap between Evidence and Policy for Infectious Diseases: How Models Can Aid Public Health Decision-Making." *Int J Infect Dis.* 42 (2016): 17–23.

Kodama, Yuichi, Jun Mashima, Takehide Kosuge, Toshiaki Katayama, Takatomo Fujisawa, Eli Kaminuma, Osamu Ogasawara, Kousaku Okubo, Toshihisa Takagi, and Yasukazu Nakamura. "The DDBJ Japanese Genotype-Phenotype Archive for Genetic and Phenotypic Human Data." *Nucleic Acids Research* 43, no. Database issue (January 28, 2015): D18–22. https://doi.org/10.1093/nar/gku1120.

Kodama, Yuichi, Martin Shumway, Rasko Leinonen, and International Nucleotide Sequence

Database Collaboration. “The Sequence Read Archive: Explosive Growth of Sequencing Data.” *Nucleic Acids Research* 40, no. Database issue (January 2012): D54-56. https://doi.org/10.1093/nar/gkr854.

Kovalsky, Anton. “COVID-19 Workspaces, Data and Tools in Terra.” COVID-19 workspaces, data and tools in Terra - Terra Support, April 16, 2020. http://support.terra.bio/hc/en-us/articles/360041068771.

Kusebauch, Ulrike, Eric W. Deutsch, David S. Campbell, Zhi Sun, Terry Farrah, and Robert L. Moritz. “Using PeptideAtlas, SRMAtlas, and PASSEL: Comprehensive Resources for Discovery and Targeted Proteomics.” *Current Protocols in Bioinformatics* 46 (June 17, 2014): 13.25.1-28. https://doi.org/10.1002/0471250953.bi1325s46.

Lapp, Hilmar. “Minimum Information About a Phylogenetic Analysis.” GitHub, May 9, 2017. https://github.com/evoinfo/miapa.

Lappalainen, Ilkka, Jeff Almeida-King, Vasudev Kumanduri, Alexander Senf, John Dylan Spalding, Saif ur-Rehman, Gary Saunders, et al. “The European Genome-Phenome Archive of Human Data Consented for Biomedical Research.” *Nature Genetics* 47, no. 7 (July 1, 2015): 692–95. https://doi.org/10.1038/ng.3312.

Lawson, Catherine L., Matthew L. Baker, Christoph Best, Chunxiao Bi, Matthew Dougherty, Powei Feng, Glen van Ginkel, et al. “EMDataBank.Org: Unified Data Resource for CryoEM.” *Nucleic Acids Research* 39, no. Database issue (January 2011): D456-464. https://doi.org/10.1093/nar/gkq880.

Lawson, Catherine L., Helen M. Berman, and Wah Chiu. “Evolving Data Standards for Cryo-EM Structures.” *Structural Dynamics* 7, no. 1 (January 1, 2020): 014701. https://doi.org/10.1063/1.5138589.

Lee, Jamie A., Josef Spidlen, Keith Boyce, Jennifer Cai, Nicholas Crosbie, Mark Dalphin, Jeff Furlong, et al. “MIFlowCyt: The Minimum Information about a Flow Cytometry Experiment.” *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* 73, no. 10 (October 2008): 926–30. https://doi.org/10.1002/cyto.a.20623.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics (Oxford, England)* 25, no. 16 (August 15, 2009): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Library of Congress. “Recommended Formats Statement – Table of Contents | Resources (Preservation, Library of Congress).” Web page. Accessed May 10, 2020. https://www.loc.gov/preservation/resources/rfs/TOC.html.

LSRI. “LSRI Response to COVID-19.” European Life Science Research Infrastructure, 2020. https://lifescience-ri.eu/ls-ri-response-to-covid-19.html.

Ma, Jie, Tao Chen, Songfeng Wu, Chunyuan Yang, Mingze Bai, Kunxian Shu, Kenli Li, et al. “IProX: An Integrated Proteome Resource.” *Nucleic Acids Research* 47, no. Database issue (January 8, 2019): D1211–17. https://doi.org/10.1093/nar/gky869.

Maddison, David R., David L. Swofford, and Wayne P. Maddison. “Nexus: An Extensible File Format for Systematic Information.” *Systematic Biology* 46, no. 4 (December 1, 1997): 590–621. https://doi.org/10.1093/sysbio/46.4.590.

Mailman, Matthew D., Michael Feolo, Yumi Jin, Masato Kimura, Kimberly Tryka, Rinat Bagoutdinov, Luning Hao, et al. “The NCBI DbGaP Database of Genotypes and Phenotypes.” *Nature Genetics* 39, no. 10 (October 2007): 1181–86.

https://doi.org/10.1038/ng1007-1181.

Majovski, Robert. “Broad Scientists Release COVID-19 Best-Practices Workflows and Analysis Tools in Terra.” Terra Support, April 16, 2020. http://support.terra.bio/hc/en-us/articles/360040613432.

Martinez-Martin, Nicole, and David Magnus. “Privacy and Ethical Challenges in Next-Generation Sequencing.” *Expert Review of Precision Medicine and Drug Development* 4, no. 2 (March 4, 2019): 95–104. https://doi.org/10.1080/23808993.2019.1599685.

Michel-Sendis, Franco. “Joint Evaluated Fission and Fusion (JEFF) Nuclear Data Library.” JEFF Nuclear Data Library - NEA, November 2017. https://www.oecd-nea.org/dbdata/jeff/.

MPEG, the Moving Picture Experts Group., and ISO/IEC JTC1/SC29/WG11. “White Paper on the Objectives and Benefits of the MPEG-G Standard.” MPEG, 2018. https://mpeg.chiariglione.org/sites/default/files/files/standards/docs/w15047-v2-w15047_GenomeCompressionStorage.zip.

National Genomics Data Center. “2019nCovR - China National Center for Bioinformation,” 2020. https://bigd.big.ac.cn/ncov?lang=en.

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

NCBI. “BioSample Database.” Home - BioSample - NCBI, 2013. https://www.ncbi.nlm.nih.gov/biosample/.

———. “BLAST TOPICS.” BLAST TOPICS, 2013. https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=BlastHelp.

———. “Gene Expression Omnibus (GEO).” Home - GEO - NCBI, 2002. https://www.ncbi.nlm.nih.gov/geo/.

———. “Sequence Read Archive (SRA).” Home - SRA - NCBI, October 3, 2019. https://www.ncbi.nlm.nih.gov/sra/.

———. “Viral Genomes.” Viral Genomes, 2020. https://www.ncbi.nlm.nih.gov/genome/viruses/.

nestor. “Nestor - Seal for Trustworthy Digital Archives.” Accessed May 10, 2020. https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html.

Nextstrain Team, Trevor Bedford, and Richard Neher. “Nextstrain Genomic epidemiology of novel coronavirus,” 2020. https://nextstrain.org/ncov/global.

NHS. “NHS Digital Leading the Protection of Patient Data with New Patient De-Identification Solution.” NHS Digital: News, August 31, 2018. https://digital.nhs.uk/news-and-events/latest-news/nhs-digital-leading-the-protection-of-patient-data-with-new-patient-de-identification-solution.

———. “The Caldicott Principles.” Information Governance Toolkit, 2013. https://www.igt.hscic.gov.uk/Caldicott2Principles.aspx.

NIH. “ClinicalTrials - Listed Clinical Studies Related to the Coronavirus Disease (COVID-19).” U.S. National Institutes of Health - Information on Clinical Trials and Human Research Studies - National Library of Medicine, 2020. https://clinicaltrials.gov/ct2/results?cond=COVID-19.

———. “Open-Access Data and Computational Resources to Address COVID-19.” National

Institutes of Health, U.S. Department of Health and Human Services, 2020. https://datascience.nih.gov/covid-19-open-access-resources.

———. "The Trans-NIH BioMedical Informatics Coordinating Committee (BMIC)." Product, Program, and Project Descriptions. National Institutes of Health, U.S. Department of Health and Human Services. U.S. National Library of Medicine, 2018. https://www.nlm.nih.gov/NIHbmic/index.html.

NIH-NCBI. "NCBI Virus: Severe Acute Respiratory Syndrome-Related Coronavirus, Taxid:694009." National Institutes of Health - National Center for Biotechnology Information, 2020. https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Nucleotide&VirusLineage_ss=Severe%20acute%20respiratory%20syndrome-related%20coronavirus,%20taxid:694009.

———. "NCBI Virus: Submit Sequences." National Institutes of Health - National Center for Biotechnology Information, 2020. https://www.ncbi.nlm.nih.gov/labs/virus/vssi/docs/submit/.

———. "Sequence Read Archive (SRA) Submission Quick Start." National Institutes of Health - National Center for Biotechnology Information, 2020. https://www.ncbi.nlm.nih.gov/sra/docs/submit/.

Nishiura, Hiroshi. "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," 2017. https://www.jst.go.jp/ristex/stipolicy/en/project/project20.html.

Ó Cathaoir, Katherina, Eugenijus Gefenas, Mette Hartlev, Miranda Mourby, and Vilma Lukaseviciene. "A European Standardization Framework for Data Integration and Data-Driven in Silico Models for Personalized Medicine – EU-STANDS4PM," March 2020. https://www.eu-stands4pm.eu/lw_resource/datapool/systemfiles/cbox/329/live/lw_datei/wp3_march2020_d3-1_v1_public.pdf.

O'Donnell, Valerie, Michael Wakelam, Shankar Subramaniam, and Ed Dennis. "LIPIDMAPS," 2020. http://www.lipidmaps.org.

OECD. "OECD Privacy Principles," August 9, 2010. http://oecdprivacy.org/.

Ogasawara, Osamu, Yuichi Kodama, Jun Mashima, Takehide Kosuge, and Takatomo Fujisawa. "DDBJ Database Updates and Computational Infrastructure Enhancement." *Nucleic Acids Research* 48, no. D1 (January 8, 2020): D45–50. https://doi.org/10.1093/nar/gkz982.

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

Ohmann, Christian. "Sharing and Reuse of Individual Participant Data from Clinical Trials: Principles and Recommendations." *6 October 2017* 7 (2017): e018647. https://doi.org/10.1136/ bmjopen-2017-018647.

Okuda, Shujiro, Yu Watanabe, Yuki Moriya, Shin Kawano, Tadashi Yamamoto, Masaki Matsumoto, Tomoyo Takami, et al. "JPOSTrepo: An International Standard Data Repository for Proteomes." *Nucleic Acids Research* 45, no. D1 (January 4, 2017): D1107–11. https://doi.org/10.1093/nar/gkw1080.

Olson, Gary. "Interpretation of the 'Newick's 8:45' Tree Format Standard." "Newick's 8:45"

Tree Format Standard, August 30, 1990. http://evolution.genetics.washington.edu/phylip/newick_doc.html.

OpenAIRE. "ARGOS Data Management Plans Creator." Accessed May 14, 2020. https://argos.openaire.eu/home.

OPIDoR. "DMP OPIDoR." Accessed May 14, 2020. https://dmp.opidor.fr/.

Oxford University. "COVID19 Dataset." Dataset, 2020. https://github.com/owid/covid-19-data.

———. "COVID19 Government Response Tracker." Dataset. University of Oxford, 2020. https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker.

Parra-Calderón, Carlos Luis, Jane Kaye, Alberto Moreno-Conde, Harriet Teare, and Francisco Nuñez-Benjumea. "Desiderata for Digital Consent in Genomic Research." *Journal of Community Genetics* 9, no. 2 (2018): 191–94. https://doi.org/10.1007/s12687-017-0355-z.

Pathak, Elizabeth Barnett, Jason L. Salemi, Natasha Sobers, Janelle Menard, and Ian R. Hambleton. "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 (April 16, 2020). https://doi.org/10.1097/PHH.0000000000001190.

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

PDBe-KB consortium. "PDBe-KB: A Community-Driven Resource for Structural and Functional Annotations." *Nucleic Acids Research* 48, no. D1 (08 2020): D344–53. https://doi.org/10.1093/nar/gkz853.

Pearson, W. R., and D. J. Lipman. "Improved Tools for Biological Sequence Comparison." *Proceedings of the National Academy of Sciences of the United States of America* 85, no. 8 (April 1988): 2444–48. https://doi.org/10.1073/pnas.85.8.2444.

Perez-Riverol, Yasset, Attila Csordas, Jingwen Bai, Manuel Bernal-Llinares, Suresh Hewapathirana, Deepti J Kundu, Avinash Inuganti, et al. "The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data." *Nucleic Acids Research* 47, no. D1 (January 8, 2019): D442–50. https://doi.org/10.1093/nar/gky1106.

Phillips, Mark, and Bartha M Knoppers. "The Discombobulation of De-Identification." *Nature Biotechnology* 34, no. 11 (November 8, 2016): 1102–3. https://doi.org/10.1038/nbt.3696.

Portage. "Assistant PGD – Réseau Portage." Accessed May 14, 2020. https://assistant.portagenetwork.ca/.

PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd. "ISO 16363." PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd. Accessed May 10, 2020. http://www.iso16363.org/.

Pupier, Marion, Jean-Marc Nuzillard, Julien Wist, Nils E. Schlörer, Stefan Kuhn, Mate Erdelyi, Christoph Steinbeck, et al. "NMReDATA, a Standard to Report the NMR Assignment and Parameters of Organic Compounds." *Magnetic Resonance in Chemistry* 56, no. 8 (2018): 703–15. https://doi.org/10.1002/mrc.4737.

Rambaut, Andrew. "Phylogenetic Analysis of NCoV-2019 Genomes." Edinburgh UK: University of Edinburgh, March 6, 2020. http://virological.org/t/phylodynamicanalysis-176-genomes-6-mar-2020/356.

———. “Virological: Novel 2019 Coronavirus Discussion Forum.” Virological, 2020. http://virological.org/c/novel-2019-coronavirus.

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

RDA-CODATA Legal Interoperability Interest Group. “Legal Interoperability of Research Data: Principles and Implementation Guidelines.” Zenodo, October 20, 2016. https://doi.org/10.5281/zenodo.162241.

RDA-COVID19-Omics Subgroup. “RDA-COVID19-Omics.” RDA, March 30, 2020. https://www.rd-alliance.org/groups/rda-covid19-omics.

Renieri, Alessandra. “GEN-COVID: Impact of Host Genome on COVID-19 Clinical Variability.” GEN-COVID, 2020. https://sites.google.com/dbm.unisi.it/gen-covid.

Research Data Alliance International Indigenous Data Sovereignty Interest Group. “CARE Principles of Indigenous Data Governance.” Global Indigenous Data Alliance, 2019. https://static1.squarespace.com/static/5d3799de845604000199cd24/t/5da9f4479ecab221ce848fb2/1571419335217/CARE+Principles_One+Pagers+FINAL_Oct_17_2019.pdf.

Ritchie, Felix. “Secure Access to Confidential Microdata: Four Years of the Virtual Microdata Laboratory.” *Economic & Market Labour Review* 2, no. 5 (May 2008): 29–34.

Rubelt, Florian, Christian E. Busse, Syed Ahmad Chan Bukhari, Jean-Philippe Bürckert, Encarnita Mariotti-Ferrandiz, Lindsay G. Cowell, Corey T. Watson, et al. “Adaptive Immune Receptor Repertoire Community Recommendations for Sharing Immune-Repertoire Sequencing Data.” *Nature Immunology* 18, no. 12 (November 16, 2017): 1274–78. https://doi.org/10.1038/ni.3873.

Rynearson, Shawn. “GFF and GVF Specification Documents.” The-Sequence-Ontology/Specifications, November 12, 2019. https://github.com/The-Sequence-Ontology/Specifications.

Sansone, Susanna-Assunta, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, Chris Taylor, Oliver Hofmann, Hong Fang, et al. “Toward Interoperable Bioscience Data.” *Nature Genetics* 44, no. 2 (February 2012): 121–26. https://doi.org/10.1038/ng.1054.

Saulnier, Katie M., David Bujold, Stephanie O. M. Dyke, Charles Dupras, Stephan Beck, Guillaume Bourque, and Yann Joly. “Benefits and Barriers in the Design of Harmonized Access Agreements for International Data Sharing.” *Scientific Data* 6, no. 1 (December 2019): 297. https://doi.org/10.1038/s41597-019-0310-4.

Schroeder, Doris. “A Global Ethics Code to Fight ‘ethics Dumping’ in Research,” 2020. https://www.globalcodeofconduct.org/.

Semantic Scholar. “CORD-19,” 2020. https://pages.semanticscholar.org/coronavirus-research.

Shanghai Public Health Clinical Center & School of Public Health. *Severe Acute Respiratory Syndrome Coronavirus 2 Isolate Wuhan-Hu-1, Complete Genome* (version MN908947.3). GenBank. Shanghai, China: Shanghai Public Health Clinical Center & School of Public Health, 2020. http://www.ncbi.nlm.nih.gov/nuccore/MN908947.3.

Sharma, Vagisha, Josh Eckels, Birgit Schilling, Christina Ludwig, Jacob D. Jaffe, Michael J. MacCoss, and Brendan MacLean. “Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline.” *Molecular & Cellular Proteomics* 17, no. 6

(June 1, 2018): 1239–44. https://doi.org/10.1074/mcp.RA117.000543.

Sharma, Vagisha, Josh Eckels, Greg K. Taylor, Nicholas J. Shulman, Andrew B. Stergachis, Shannon A. Joyner, Ping Yan, et al. "Panorama: A Targeted Proteomics Knowledge Base." *Journal of Proteome Research* 13, no. 9 (September 5, 2014): 4205–10. https://doi.org/10.1021/pr5006636.

Spidlen, Josef, Ryan Brinkman, and ISAC Data Standards Task Force. "Gating-ML 2.0." FAIRsharing, March 16, 2015. https://doi.org/10.25504/FAIRSHARING.QPYP5G.

Spidlen, Josef, Wayne Moore, ISAC Data Standards Task Force, and Ryan R. Brinkman. "ISAC's Gating-ML 2.0 Data Exchange Standard for Gating Description." *Cytometry Part A* 87, no. 7 (July 2015): 683–87. https://doi.org/10.1002/cyto.a.22690.

Spidlen, Josef, Wayne Moore, David Parks, Michael Goldberg, Chris Bray, Pierre Bierre, Peter Gorombey, et al. "Data File Standard for Flow Cytometry, Version FCS 3.1." *Cytometry Part A* 77, no. 1 (January 2010): 97–100. https://doi.org/10.1002/cyto.a.20825.

Stoltzfus, Arlin, Brian O'Meara, Jamie Whitacre, Ross Mounce, Emily L Gillespie, Sudhir Kumar, Dan F Rosauer, and Rutger A Vos. "Sharing and Re-Use of Phylogenetic Trees (and Associated Data) to Facilitate Synthesis." *BMC Research Notes* 5, no. 1 (December 2012): 574. https://doi.org/10.1186/1756-0500-5-574.

Sud, Manish, Eoin Fahy, Dawn Cotter, Kenan Azam, Ilango Vadivelu, Charles Burant, Arthur Edison, et al. "Metabolomics Workbench: An International Repository for Metabolomics Data and Metadata, Metabolite Standards, Protocols, Tutorials and Training, and Analysis Tools." *Nucleic Acids Research* 44, no. Database issue (January 4, 2016): D463–70. https://doi.org/10.1093/nar/gkv1042.

Swiss Institute of Bioinformatics (SIB). "SARS-COV-2, COVID-19 Coronavirus Resource: SARS Coronavirus 2 (SARS-CoV-2) Proteome." SARS coronavirus 2 ~ ViralZone page, 2020. https://viralzone.expasy.org/8996.

Taylor, Chris F, Norman W Paton, Kathryn S Lilley, Pierre-Alain Binz, Randall K Julian, Andrew R Jones, Weimin Zhu, et al. "The Minimum Information about a Proteomics Experiment (MIAPE)." *Nature Biotechnology* 25, no. 8 (August 2007): 887–93. https://doi.org/10.1038/nbt1329.

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, 2017. 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." ISO, 2013. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/58/75852.html.

Templ, Matthias, Bernhard Meindl, and Alexander Kowarik. *SdcMicro: Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation*, 2020. https://cran.r-project.org/web/packages/sdcMicro/index.html.

The Atlantic. "COVID Tracking Project." Dataset, 2020. https://covidtracking.com/.

The Human Protein Atlas consortium. "SARS-CoV-2 Related Proteins - The Human Protein Atlas," 2020. https://www.proteinatlas.org/humanproteome/sars-cov-2.

The ImmPort project. "ImmPort Shared Data." ImmPort, 2018. https://immport.org/shared/home.

The Open Covid Pledge. "The Open Covid Pledge," April 7, 2020. https://opencovidpledge.org/.

The South African San Institute. "The San Code of Research Ethics," 2017. http://trust-project.eu/wp-content/uploads/2017/03/San-Code-of-RESEARCH-Ethics-Booklet-final.pdf.

The United Nations. "The UN Ethics Office." The UN Ethics Office: Listen - Advise - Respect, 2020. https://www.un.org/en/ethics/index.shtml.

UCSC Genome Bioinformatics group. "Genome Browser FAQ: BED (Browser Extensible Data) Format." UCSC Genome Browser, 2020. http://genome.ucsc.edu/FAQ/FAQformat#format1.

UCSF Computer Graphics Laboratory. "Aligned FASTA Format," November 2009. https://www.cgl.ucsf.edu/chimera/docs/ContributedSoftware/multalignviewer/afasta.html.

UK Data Service. "Recommended Formats." Accessed May 10, 2020. https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats.

———. "Regulating Access to Data," 2020. https://www.ukdataservice.ac.uk/manage-data/legal-ethical/access-control/five-safes.

Ulrich, Eldon L. *NMR-STAR Dictionary* (version 3.2.1.20), 2019. https://github.com/uwbmrb/nmr-star-dictionary.

Ulrich, Eldon L., Hideo Akutsu, Jurgen F. Doreleijers, Yoko Harano, Yannis E. Ioannidis, Jundong Lin, Miron Livny, et al. "BioMagResBank." *Nucleic Acids Research* 36, no. suppl_1 (January 1, 2008): D402–8. https://doi.org/10.1093/nar/gkm957.

———. "BMRB - Biological Magnetic Resonance Bank." BMRB - Biological Magnetic Resonance Bank, 2008. http://www.bmrb.wisc.edu/.

Ulrich, Eldon L., Kumaran Baskaran, Hesam Dashti, Yannis E. Ioannidis, Miron Livny, Pedro R. Romero, Dimitri Maziuk, et al. "NMR-STAR: Comprehensive Ontology for Representing, Archiving and Exchanging Data from Nuclear Magnetic Resonance Spectroscopic Experiments." *Journal of Biomolecular NMR* 73, no. 1 (February 1, 2019): 5–9. https://doi.org/10.1007/s10858-018-0220-3.

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

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

UN Office of the High Commissioner. "COMMITTEE ON ECONOMIC, SOCIAL AND CULTURAL RIGHTS," 2020. https://www.ohchr.org/en/hrbodies/cescr/pages/cescrindex.aspx.

UNDRR. "Disaster Risk Management for Health: Overview," 2020. https://www.undrr.org/publication/disaster-risk-management-health-overview.

UNESCO. "Universal Declaration on Bioethics and Human Rights," October 19, 2005. https://en.unesco.org/themes/ethics-science-and-technology/bioethics-and-human-rights.

UNESCO International Bioethics Committee. "Report of the IBC on the Principle of the Sharing of Benefits," October 15, 2015.

https://unesdoc.unesco.org/ark:/48223/pf0000233230.

UNESCO International Bioethics Committee, and UNESCO World Committion on the Ethics of Scientific Knowledge and Technology. "STATEMENT ON COVID-19: ETHICAL CONSIDERATIONS FROM A GLOBAL PERSPECTIVE," 2020. https://unesdoc.unesco.org/ark:/48223/pf0000373115.

Unidata Program center, UCAR. "The NetCDF-C Tutorial: The NetCDF Data Model." NetCDF: The NetCDF Data Model, March 27, 2020. https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_model.html.

UniProt. "COVID-19 UniProtKB." UniProt, 2020. https://covid-19.uniprot.org/uniprotkb?query=*.

University of California. "DMP Tool." University of California Curation Center, 2016. https://dmptool.org/.

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

University of Washington. "COVID19 Data: Beoutbreakprepared," 2020. https://github.com/beoutbreakprepared.

Vander Heiden, Jason Anthony, Susanna Marquez, Nishanth Marthandan, Syed Ahmad Chan Bukhari, Christian E. Busse, Brian Corrie, Uri Hershberg, et al. "AIRR Community Standardized Representations for Annotated Immune Repertoires." *Frontiers in Immunology* 9 (September 28, 2018): 2206. https://doi.org/10.3389/fimmu.2018.02206.

Vilches, Claudia. "Biblioguias: Gestión de Datos de Investigación: Módulo 2 - Plan de Gestión de Datos (PGD)." Accessed May 14, 2020. https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/PGD.

VIVLI. "Center for Global Clinical Research Data," 2020. https://vivli.org/.

Wang, Mingxun, Jian Wang, Jeremy Carver, Benjamin S. Pullman, Seong Won Cha, and Nuno Bandeira. "Assembling the Community-Scale Discoverable Human Proteome." *Cell Systems* 7, no. 4 (October 24, 2018): 412-421.e5. https://doi.org/10.1016/j.cels.2018.08.004.

White House. "COVID-19 Open Research Dataset Challenge (CORD-19)," 2020. https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge.

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

———. "Ethical Considerations in Developing a Public Health Response to Pandemic Influenza," 2007. https://apps.who.int/iris/bitstream/handle/10665/70006/WHO_CDS_EPR_GIP_2007.2_eng.pdf.

———. "Global Surveillance for COVID-19 Caused by Human Infection with COVID-19 Virus: Interim Guidance." World Health Organization, March 20, 2020. https://apps.who.int/iris/bitstream/handle/10665/331506/WHO-2019-nCoV-SurveillanceGuidance-2020.6-eng.pdf.

———. "Modes of Transmission of Virus Causing COVID-19: Implications for IPC Precaution Recommendations," April 29, 2020. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations.

———. "Novel Coronavirus (2019-NCoV) Situation Reports." World Health Organization, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

———. “Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19).” World Health Organization, February 28, 2020. https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19).

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

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3, no. 1 (2016): 1–9. https://doi.org/10.1038/sdata.2016.18.

Willenborg, Leon, and Ton de Waal. *Element of Statistical Disclosure Control*. Lecture Notes in Statistics 155. New York: Springer, 2001.

World Bank. “Understanding the Coronavirus (COVID-19) Pandemic through Data.” Datasets, 2020. http://datatopics.worldbank.org/universal-health-coverage/covid19/.

Worldometers. “COVID19 Data.” Dataset, 2020. https://www.worldometers.info/coronavirus/.

Worldwide Protein Data Bank(wwPDB). “PDBx/MmCIF Dictionary Resources.” PDBx/mmCIF Dictionary Resources, 2014. http://mmcif.pdb.org/.

wwPDB Consortium. *WwPDB OneDep System* (version 4.5). wwPDB Consortium, 2020. https://deposit-2.wwpdb.org/.

wwPDB consortium, Stephen K Burley, Helen M Berman, Charmi Bhikadiya, Chunxiao Bi, Li Chen, Luigi Di Costanzo, et al. “Protein Data Bank: The Single Global Archive for 3D Macromolecular Structure Data.” *Nucleic Acids Research* 47, no. D1 (January 8, 2019): D520–28. https://doi.org/10.1093/nar/gky949.

Xu, Bo, Bernardo Gutierrez, Sumiko Mekaru, Kara Sewalk, Lauren Goodwin, Alyssa Loskill, Emily L. Cohn, et al. “Epidemiological Data from the COVID-19 Outbreak, Real-Time Case Information.” *Scientific Data* 7, no. 1 (December 2020): 106. https://doi.org/10.1038/s41597-020-0448-0.

Yilmaz, Pelin, Renzo Kottmann, Dawn Field, Rob Knight, James R. Cole, Linda Amaral-Zettler, Jack A. Gilbert, et al. “Minimum Information about a Marker Gene Sequence (MIMARKS) and Minimum Information about Any (x) Sequence (MIxS) Specifications.” *Nature Biotechnology* 29, no. 5 (May 2011): 415–20. https://doi.org/10.1038/nbt.1823.

Zhang, Kai-Yue, Yi-Zhou Gao, Meng-Ze Du, Shuo Liu, Chuan Dong, and Feng-Biao Guo. “Vgas: A Viral Genome Annotation System.” *Frontiers in Microbiology* 10 (February 13, 2019). https://doi.org/10.3389/fmicb.2019.00184.

Zhang, Yanping. “The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) — China, 2020.” Cina CDC Weekly, February 17, 2020. http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51.

12. Contributors

We would like to acknowledge the global cohort of RDA community members who have contributed their time, knowledge and expertise to generate these guidelines. Listed below according to: First Name Last Name (ORCID)

Clara Amid 0000-0001-6534-7425
Pamela Andanda 0000-0002-2746-7861
Claire Austin 0000-0001-9138-5986 C.
Christophe Bahim
Michelle Barker 0000-0002-3623-172X
Marlon Bayot 0000-0002-5328-150X
Alexandre Beaufays
Alexander Bernier 0000-0001-8615-8375
Louise Bezuidenhout 0000-0003-4328-3963
Juan Bicarregui 0000-0001-5250-7653
Timea Biro
Hélène Blasco 0000-0001-6107-0035
Sabrina Boni
Sergio Bonini
Christian Busse 0000-0001-7553-905X
Korbinian Bösl 0000-0003-0498-4273
Anne Cambon-Thomsen 0000-0001-8793-3644
Stephanie Carroll 0000-0002-8996-8071
Leyla Castro Jael Garcia 0000-0003-3986-0510
Calvin Chan Wing Yiu 0000-0002-3656-7709
Jorge Clarke 0000-0003-1314-7020
Brian Corrie 0000-0003-3888-6495
Zoe Cournia 000-0001-9287-264X
Piotr Dabrowski Wojciech 0000-0003-4893-805X
Luc Decker 0000-0002-4808-3568
Laurence Delhaes 0000-0001-7489-9205
David Delmail
Cyrille Delpierre
Natalie Dewson 0000-0002-5968-9696
Kheeran Dharmawardena
Gayo Diallo
Ingrid Dillo 0000-0001-5654-2392
Diana Dimitrova 0000-0003-4732-7054
Stephan Druskat 0000-0003-4925-7248
Thomas Duflot 0000-0002-8730-284X
Patrick Dunn 0000-0003-1868-9689
Nora Dörrenbächer 0000-0002-6246-1051
Claudia Engelhardt 0000-0002-3391-7638
Keyvan Farahani 0000-0003-2111-1896
Juliane Fluck 0000-0003-1379-7023
Konrad Förstner 0000-0002-1481-2996
Leyla Garcia 0000-0003-3986-0510
Sandra Gesing 0000-0002-6051-0673
Carole Goble 0000-0003-1219-2137
Martin Golebiewski 0000-0002-8683-7084
Alejandra Gonzalez-Beltran 0000-0003-3499-8262
Jay Greenfield
Wei Gu 0000-0003-3951-6680
Anupama Gururaj 0000-0002-4221-4379
Dara Hallinan 0000-0002-1160-821X
Natalie Harrower 0000-0002-7487-4881
Pascal Heus 0000-0002-6543-7102
Pieter Heyvaert 0000-0002-1583-5719
Neil Hong Chue 0000-0002-8876-7606
Rob Hooft 0000-0001-6825-9439
Wim Hugo 0000-0002-0255-5101
Andrea Jackson-Dipina
Ann James Myatt 0000-0002-2137-7961
Sarah Jones
Chifundo Kanjala
Daniel Katz 0000-0001-5934-7525 S.
Iryna Kuchma 000-0002-2064-3439
Helena Laaksonen 0000-0002-1312-1958
Ann-Lena Lamprecht 0000-0003-1953-5606
Dollé Laurent 0000-0003-4566-6407
Paula Lavanchy Martinez 0000-0003-1448-0917
Young-Joo Lee 0000-0001-7189-6607
Mark Leggott
Joanna Leng 0000-0001-9790-162X
Marcia Levenstein
Dawei Lin 0000-0002-5506-0030
Birte Lindstaedt
Aliaksandra Lisouskaya 0000-0001-7556-8977
Nicolas Loozen
Paula Martinez Andrea 0000-0002-8990-1985
Gary Mazzaferro 0000-0003-2773-5317
Katherine McNeil 0000-0003-2865-3751
Claudia Bauzer Medeiros 0000-0003-1908-4753
Eva Méndez
Natalie Meyers 0000-0001-6441-6716
Robin Michelet

Daniel Mietchen
Ingvill Constanze Mochmann 0000-0002-5481-3432
David Molik 0000-0003-3192-6538
Laura Morales 0000-0002-6688-6508
Rowland Mosbergen 0000-0003-1351-8522
Rajini Nagrani 0000-0002-1708-2319
Diana Navarro-Llobet 0000-0002-0563-3937
Gustav Nilsonne 0000-0001-5273-0150
Jenny O'Neill 0000-0002-1644-1236
Christian Ohmann 0000-0002-5919-1003
Natalie Pankova 0000-0002-7218-3518
Simon Parker 0000-0001-9993-533X
Carlos Parra-Calderon Luis 0000-0003-2609-575X
Pablo de Pedraza
Pandelis Perakakis 0000-0002-9130-3247
Brian Pickering 0000-0002-6815-2938
Amy Pienta 0000-0003-1174-6118
Priyanka Pillai
Eric Piver 0000-0002-7101-0121
Panayiota Polydoratou 0000-0002-7551-8002
Fotis Psomopoulos 0000-0002-0222-4273
Rob Quick 0000-0002-0994-728X
Valeria Quochi 0000-0002-1321-5444
Dana Rad 0000-0001-6754-3585
Alessandra Renieri 0000-0002-0846-9220
Stéphanie Rennes 0000-0003-1458-7773
Artur Rocha 0000-0002-5637-1041
Susanna-Assunta Sansone 0000-0001-5306-5690
Venkata Satagopam 0000-0002-6532-5880
Stefan Sauermann
Carsten Schmidt Oliver
Meg Sears
Hugh Shanahan 0000-0003-1374-6015
Tim Smith 0000-0002-1567-7116
Joanne Stocks
Rainer Stotzka 0000-0003-3642-1264
Shoaib Sufi 0000-0001-6390-2616
Mark Taylor 0000-0003-2009-6284
Marta Teperek 0000-0001-8520-5598
Mogens Thomsen 0000-0002-4546-0129
Henri Tonnang
Valeria Quochi 0000-0002-1321-5444
Marcos Roberto Tovani-Palone 0000-0003-1149-2437
Gabriel Turinici
Yasemin Türkyilmaz-Van der Velden 0000-0003-2562-0452
Mary Uhlmansiek
Meghan Underwood
Justine Vandendorpe 0000-0002-9421-8582
Bridget Walker
Minglu Wang 0000-0002-0021-5605
Galia Weidl
Anna Widyastuti
Kara Woo 0000-0002-5125-4188
Qian Zhang 0000-0003-1549-7358