RDA COVID-19
Recommendations and Guidelines on Data Sharing

Abstract: This is the final version of the Recommendations and Guidelines from the RDA COVID19 Working Group, and has been endorsed through the official RDA process.

RDA COVID-19

Recommendations and Guidelines on Data Sharing

RDA Recommendation (FINAL Release)

Produced by: RDA COVID-19 Working Group, 2020

Document Metadata

Identifier	DOI: https://doi.org/10.15497/rda00052
Citation	To cite this document please use: RDA COVID-19 Working Group. Recommendations and Guidelines on data sharing. Research Data Alliance. 2020. DOI: https://doi.org/10.15497/rda00052
Title	RDA COVID-19; Recommendations and Guidelines on Data Sharing, Final release 30 June 2020
Description	This is the final version of the Recommendations and Guidelines from the RDA COVID-19 Working Group, and has been endorsed through the official RDA process.
Date Issued	2020-06-30
Version	Final guidelines and recommendations, 30 June 2020, endorsed version
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,' `RDA International Indigenous Data Sovereignty Interest Group,' and we acknowledge the support provided by the RDA community and structure.
Licence	This work is licensed under CC0 1.0 Universal (CC0 1.0) Public Domain Dedication.
Disclaimer	The views and opinions expressed in this document are those of the individuals identified above and in the list of Contributors at the end of the document, and do not necessarily reflect the official policy or position of their respective employers, or of any government agency or organisation.
Group Co-chairs	*Juan Bicarregui, Anne Cambon-Thomsen, Ingrid Dillo, Natalie Harrower, Sarah Jones, Mark Leggott, Priyanka Pillai*
Subgroup Moderators	***Clinical****: Sergio Bonini, Andrea Jackson-Dipina, Dawei Lin, Christian Ohmann* ***Community Participation****: Timea Biro, Kheeran Dharmawardena, Eva Méndez, Daniel Mietchen, Susanna Sansone, Joanne Stocks* ***Epidemiology****: Claire Austin, Gabriel Turinici* ***Indigenous Data****: Stephanie Russo Carroll* ***Legal and Ethical****: Alexander Bernier, John Brian Pickering* ***Omics****: Rob Hooft, Natalie Meyers* ***Social Sciences****: Iryna Kuchma, Amy Pienta* ***Software****: Michelle Barker, Fotis Psomopoulos, Hugh Shanahan*
Editorial Team	*Christophe Bahim, Alexandre Beaufays, Ingrid Dillo, Natalie Harrower, Mark Leggott, Nicolas Loozen, Robyn Nicholson, Priyanka Pillai, Mary Uhlmansiek, Meghan Underwood, Bridget Walker*

Table of Contents

Table of Figures

Table of Tables

Executive Summary

Background

Data holds the potential to drive rapid response and informed decision-making during public health emergencies. There is a need for timely and accurate collection, reporting and sharing of data within and between research communities, public health practitioners, clinicians and policymakers. Accurate and rapid availability of data will inform assessment of the severity, spread and impact of a pandemic to implement efficient and effective response strategies.

The availability of efficient information and communication technology has improved the global capacity to implement systems to share data during a pandemic. However, the harmonisation across these sophisticated yet diverse systems combined with the timeliness of accessing data across information systems are currently major roadblocks. The World Health Organization'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 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 Research Data Alliance (RDA) COVID-19 Working Group (CWG) members bring varied global expertise to develop a body of work that comprises how data from multiple disciplines inform response to a pandemic combined with guidelines and recommendations on data sharing under the present COVID-19 circumstances. This extends to research software sharing, in recognition of the key role played by software in analysing data. The work has been divided into four research areas (Clinical, Omics, Epidemiology, Social Sciences) with four cross-cutting themes (Community Participation, Indigenous Data, Legal and Ethical Considerations, Research Software), as a way to focus the conversations and provide an initial set of guidelines in a tight timeframe. The detailed guidelines are aimed to help **stakeholders follow best practices to maximise the efficiency of their work, and to act as a blueprint for future emergencies**. The recommendations in the document are aimed at **helping policymakers and funders to maximise timely, quality data sharing and appropriate responses in such health emergencies**.

The CWG addressed 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 policymaking, including an overarching framework, common tools and processes, and principles that can be embedded in research practice**. The guidelines address general aspects of data practice, for example the FAIR principles (that research outputs should be Findable, Accessible, Interoperable, and Reusable), or the adoption of research-domain community standards.

There are foundational overarching challenges and recommendations that appear across the four research sub-groups as well as the cross-cutting themes. These foundational elements are presented in the summary before the area-specific challenges, recommendations and guidelines are articulated.

Challenges

The unprecedented spread of the virus has prompted **a rapid and massive research** response with a diversity of outputs that pose a challenge to interoperability. 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. 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.

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. Furthermore, many outputs are not reusable by, or useful to, different communities if they have not been sufficiently documented and contextualised, or appropriately licensed. Correspondingly, research software is developed and maintained in an *ad hoc* fashion. Access information for the software developed for analysis is not noted consistently in papers and, if the software is available, it is often placed in arbitrary locations with no guarantee of its persistence.

Recommendations

Governments, research funders, and research or research-supporting institutions around the world must coordinate with one another, and support and promote **Open Science** through policy and investment to streamline the flow of data between local entities, and across international jurisdictions.

There are motivational barriers to making data outputs available rapidly. There is a need for **incentivising the early publication/release of data outputs and the software used to produce them** during a public health emergency. The early publication/release of data outputs and the tools used to create them should be encouraged by building trust, providing incentives for sharing data and providing appropriate governance.

There is a need to **invest** in state-of-the-art information technology (IT) and data management systems infrastructure. The investment should also be directed towards people and skills to fully utilise the potential of large-scale infrastructure. The **minimum required infrastructure for pandemic response** in terms of technology, skills, people and frameworks should be accessible to all jurisdictions/sectors.

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. A balance between achieving 'perfectly' **FAIR outputs** and timely sharing is necessary with the key goal of immediate and open sharing as a driver. Data management plans (DMPs) should be created early in the research process and updated regularly to prepare for data deposit and reuse.

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.

Research outputs need to be documented, which includes **documentation** of methodologies used to collect, define and construct data, data cleaning, data imputation, data provenance and so on. 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.

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. By providing persistent identifiers, requiring preferred formats, rich metadata, etc., certified trustworthy repositories already guarantee a baseline FAIRness of and sustained access to the data, as well as citation.

Pre-print journals should undergo **an expedited review process** to balance the need to publish findings rapidly with the requirement to publish relevant and reliable findings. Full reports should be made available immediately upon communication of results, e.g. through a press release. Peer-reviewed data articles should be treated as first-class research outputs equal in value to traditional peer-reviewed articles. In order to **expedite reuse**, 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.

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.

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 related legislations** activated during a pandemic need to clearly outline data custodianship/ownership, publication rights and arrangements, consent models, and permissions around sharing data and exemptions.

The sub-groups and cross-cutting themes have each articulated the challenges facing researchers working on COVID-19, as well as recommendations/guidelines for improving data sharing (Table 1). These sub-group guidelines and recommendations should be considered directly depending on the relevant area of COVID-19 research as well as policy/decision-making.

Table 1 - Summary of challenges, guidelines and recommendations

Sub-groups/cross cutting themes	Challenges	Guidelines for researchers	Recommendations for funders/policymakers
Clinical	Promotion of clinical data sharing is important due to many studies and trials being performed under enormous time pressure	Standardised clinical terminologies should be used and a fair balance achieved between timely data sharing and protecting privacy and confidentiality	Measures should be taken in order to organise the sharing of data and trial documents in a suitable, trustworthy and secure data repository

Omics	An increased need of rapid openness for omics data to gain early insights into molecular biology of the processes at cellular level	Omics research should be a collaborative effort to learn the genetic determinants of COVID-19 susceptibility, severity and outcomes	Promote use of domain-specific repositories to enable standardisation of terms and enforce metadata standards
Epidemiology	Data and models are frequently incomplete, provisional, and subject to correction under changing conditions	Data models must include clinical data, disease milestones, indicators and reporting data, contact tracing and personal risk factors	Incentivise the publication of situational data, analytical models, scientific findings, and reports used in decision-making
Social Sciences	Require equal inclusion of social and economic context with health-related information to enable evidence-based decision-making	Enable interoperable cross-disciplinary and cross-cultural data collection, data use and collaboration for managing social sciences data during pandemics	Ensure robust funding streams for social sciences research for understanding and managing the human aspects of pandemics
Community	Need specific guidelines for enabling citizen scientists undertaking research to contribute to a common body of knowledge	Encourage public and patient involvement (PPI) throughout the data management lifecycle from research question to final data sharing and usage	Balance between timely testing and contact tracing, emergency response, community safety and individual privacy concerns
Indigenous Data Guidelines	Indigenous data rights, priorities and interests must be recognised in all COVID-19 research and surveillance activities	Indigenous governance of data collection, ownership, sharing and use priorities is the central principle of Indigenous data sovereignty	CARE Principles of Indigenous Data Governance set minimum standards for collectors, users and stewards of data
Legal and Ethical Considerations	Achieve a balance between rights of people and interests of researchers and policymakers	Ethical instruments should be interpreted with the law, and can guide the interpretation of the law if the law does not address a particular issue	During a pandemic, ethical review and approval for legally sharing data should be expedited
Research Software	Need systems in place for sharing of research software and accelerated and reproducible research during a pandemic	It is critical for software that is used in data analysis to produce results that can, if necessary, be reproduced	Funders must allocate financial resources to support the development and maintenance of new research software

1. Objectives and Use of This Document

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

1. to clearly define detailed *guidelines* on data and software sharing under the present COVID-19 circumstances to help stakeholders follow best practices to maximise 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 and software sharing and appropriate responses in such health emergencies;

3. to address the interests of researchers, policymakers, funders, publishers, and providers of data sharing infrastructures.

It is important to note that in this document the terms *guidelines* and *recommendations* are distinguished as follows. A guideline provides detailed advice pertaining to the practice of research data and software sharing. As a consequence, guidelines are aimed at researchers, data stewards, research software engineers and public health officials. A recommendation provides higher level and more generic advice. As a consequence, recommendations are aimed at other important stakeholder groups such as policymakers, funders, publishers and infrastructure providers.
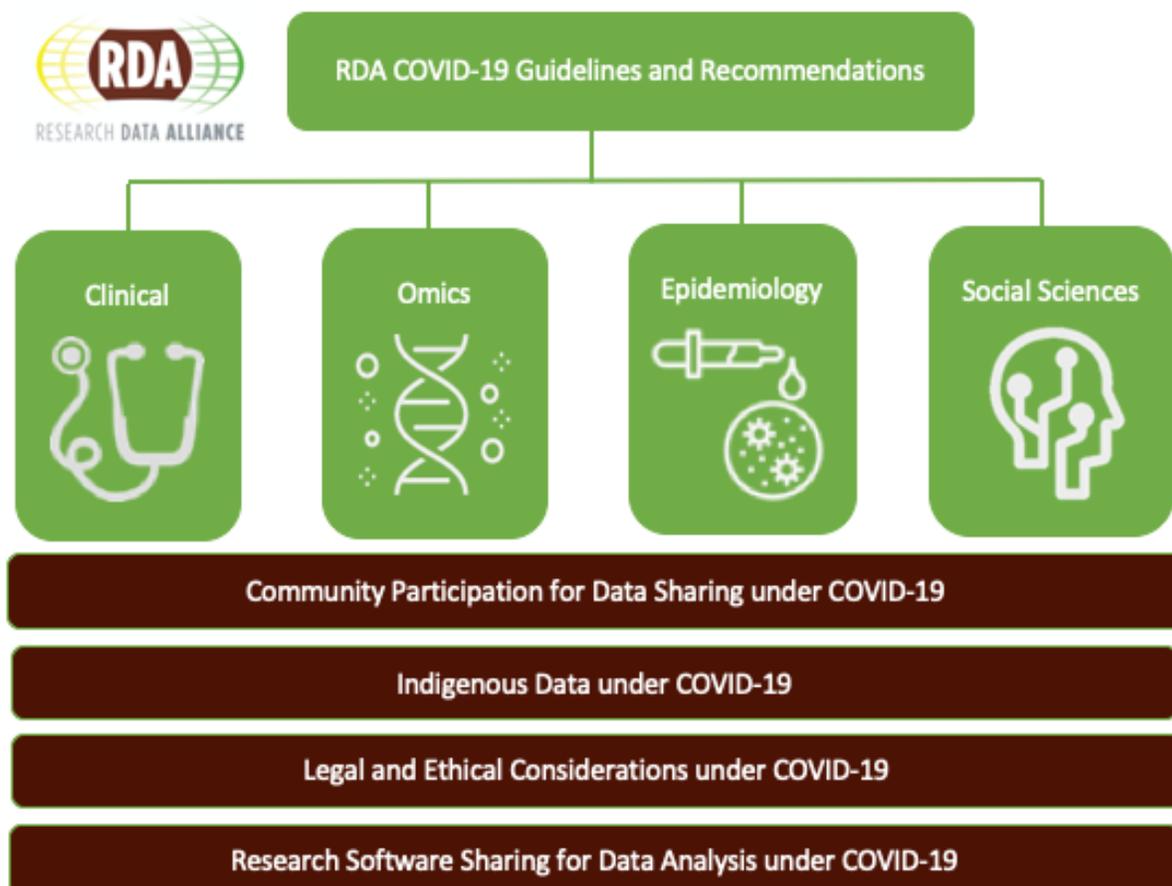
Figure 1 - RDA COVID-19 WG sub-groups including research areas and cross-cutting themes

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

The CWG is addressing the development of 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 policymaking, including an overarching framework, common tools and processes, and principles that can be embedded in research practice. The guidelines contained herein address general aspects that data should adhere to, for example the FAIR principles (that research outputs should be Findable, Accessible, Interoperable, and Reusable), or the adoption of research domain community standards. At the same time, they also provide 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.

These detailed guidelines are supplemented with higher level *recommendations* aimed at the other stakeholder groups who need to work together with researchers, data stewards and research software engineers to realise the timely and open sharing of research data and software as a key component of pandemic preparedness and response.

The work has been divided into four research areas with four 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 and the European Commission. The first meeting of the CWG to determine the work was held in March 2020. As of June 2020, the CWG counted over 440 members spread across the different sub-groups. This effort also reflects the work of a host of other RDA Working Groups, as well as external stakeholder organisations, including the Global Indigenous Data Alliance and the Research Software Alliance.

The CWG and the sub-groups operate according to the RDA guiding principles of Openness, Consensus, Balance, Harmonisation, Community-driven, Non-profit, and Technology-neutral, and are open to all.

The aims of the CWG align with recent statements made by other international organisations promoting rapid and efficient sharing of data during the pandemic. The World Health Organization'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 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. There are many initiatives developing in different communities to facilitate COVID-19 data sharing. For example, one could look to CoronaWhy - a globally distributed, volunteer-powered research community aimed at answering key questions related to COVID-19. CoronaWhy is applying recent advances in natural language processing, data mining, machine learning and other Artificial Intelligence technologies, building a horizontal data management platform following the best FAIR practices and facilitating the collaboration between different research communities from various countries.

This document starts in Section 2 with an overview of foundational, overarching elements that emerged across the different research areas. These recommendations touch upon a number of topics well known to those working in research data sharing. The focus of Sections 3 to 6 is on the COVID-19 related research areas. Each section starts with a description of the area and the focus and scope of the work, followed by the actual recommendations and guidelines. In sections 7 to 10 this same structure is used for the four cross-cutting themes. The document contains an extended glossary of terms and acronyms to support the reader (Sections 11 and 12, respectively), an overview of useful additional resources (Section 13) and a list of references (Section 14). Note as well that the RDA-COVID19 working group Zotero Library has been

published to support these Recommendations and Guidelines. Section 15 lists the contributors to this work.

The Research Data Alliance understands that this document is lengthy and may well challenge busy researchers to find the specifics that most apply to them. Please see the working group page for an Executive Summary, and visual aids for navigating the document. Please cite this document if you find it useful, as this will help to disseminate the recommendations and guidelines: RDA COVID-19 Working Group. Recommendations and Guidelines on data sharing. Research Data Alliance. 2020. DOI: https://doi.org/10.15497/rda00052.

2. Foundational Elements

Specific challenges facing researchers working on COVID-19 from different research areas have been articulated throughout this document, as well as recommendations and guidelines for improving data sharing in these areas. These recommendations and guidelines should be considered directly depending on the relevant area of COVID-19 research. However, certain common or key aspects appear across different research areas. These are presented here as foundational elements.

2.1 Challenges

The availability of research data is a key component of pandemic preparedness and response. The timeliness of accessing data and the harmonisation across information systems are currently major roadblocks.

Critical Need for Data Sharing

The unprecedented spread of the virus has prompted a rapid and massive research response. 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; due to decisions being taken by healthcare practitioners and governments on a daily basis, it is crucial 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. Shortcuts with the interpretation of data can, for example, create issues such as the early debate on the severity, transmissibility and global spread of COVID-19.

The obligation to share data could orient at least some institutions to reduce testing (only confirmed, not suspected, cases "count" and hence reducing testing allows for a lower number of confirmed cases, creating the illusion that the epidemic is under control).

And in some cases, a lack of transparency and publication of false or unchecked numbers is perhaps worse than no publication at all.

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 studies around 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 hyper speed.

Lack of Harmonised Standards and Context

Emerging infections are largely unpredictable in nature and there are 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. Furthermore, 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 they can be reused and built upon (see the OECD Open Science Policy Brief).

2.2 Recommendations

2.2.1 Coordinated, Cross-jurisdictional Efforts to Foster Global Open Science

The COVID-19 pandemic has had, in a very short time, an unprecedented global impact on health, economies, and daily life. It has underlined the importance of open science practices and demonstrated clearly how different jurisdictions require the policies and support to collaborate with and build research efforts across political, geographical, and disciplinary boundaries. In addition to sharing the response effort, effective cross-national comparisons can provide useful insights for the development of future global emergency preparedness programmes.

Governments, research funders, and research or research-supporting institutions around the world must coordinate with one another, support and promote Open Science through policy and investment to streamline the flow of data between local entities, and across international jurisdictions. Systemic investment in and support for Open Science must be developed rapidly and sustainably to face both our current pandemic and future public health emergencies.

Coordination includes such efforts as urgently updating 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 modernisation, and linked data life cycles that will enable rapid and credible scientific discovery, and fast-track decision-making. International organisations and alliances such as the World Health Organization, the OECD, the International Science Council, UNESCO and the Research Data Alliance, to name a few, provide avenues for this coordination. We also call on the international Open Government Partnership (OGP) to add "Open Science" as one of its Policy Areas (OGP, 2020a).

Similarly, governments, funders and policymakers should engage with big technology companies, mobile network operators, social network companies and others in the private sector who hold data that can better help understand the pandemic and population behaviour. Data sharing policies should be adopted to encourage and facilitate data flows from data holders to the research community with the goal of protecting citizens' rights and health (de Pedraza et al., 2019; Askitas, 2018).

There is a need for incentivising the early publication/release of data outputs during a public health emergency, because there are motivational barriers to making data outputs available rapidly and before primary papers have been written or published. Data publication should be encouraged by building trust and providing incentives and credit for preparing and sharing data and providing appropriate governance. It is important to foster collaboration under agreements that clearly describe how the data will be used

(e.g. only for early investigation of pandemic, surveillance or research and not for publication without consent and/or credit), with whom the data will be shared, and the value of sharing data for informing response during an emergency. Initiatives to support rewards and credits for data sharing should be strengthened. See RDA Sharing Rewards and Credit (SHARC) IG (Research Data Alliance) and FORCE 11 "Joint declaration of data citation principles" (Data Citation Synthesis Group, 2014). Correspondingly, research software should be made available in code repositories that enable feedback from the research community, and archival repositories which can then be referenced with persistent identifiers.

Research institutions and funding agencies can incentivise data stewardship, research software engineering data and software sharing by creating structures for researchers to get credit for this work and by providing support for publishing data and software as valid research outputs. This can include developing research assessment systems that reward data and software outputs alongside publications and other research objects. Policymakers should put guidelines into place that give researchers ease of mind when licensing/sharing their data. All research data based on public research funding should be made available and exploitable in a timely manner, in particular for those of critical interest during an emergency situation. From a funding agency perspective, this could mean that increased weighting is given in the grant review process to researchers who demonstrate best practice in open data and data reproducibility with respect to their research outputs.

2.2.2 Infrastructure Investment & Economies of Scale

Support for Open Science requires investment in infrastructure so governments, funders and institutions should therefore fund state-of-the-art information technology (IT) and data management infrastructure, which includes hardware, networks, and the development and maintenance of critical research software. Investment should also be directed towards the human resources required to maintain the infrastructure, and the training and support required to fully utilise the potential of large-scale infrastructure.

In general, new infrastructure should harness the value of existing infrastructure, building on what is already working well. Economies of scale should be considered when planning institutional, disciplinary, sector-wide, or regional/national infrastructure to reduce overlap, encourage collaboration, and maximise return on investment. In the case of limited resource settings, the use of existing data management infrastructure should be leveraged to prioritise and support pandemic research response.

Research institutions can provide granular, tiered access to restricted data by appropriately authorised and credentialed users and machines, based on roles and needs. Data storage policies should follow recommendations regarding regular backup in multiple locations. There should be priority access to resources for researchers/practitioners working on response during a public health emergency. The research community's ability to apply best practices for research software, including training in software development concepts is to be encouraged.

The investment in infrastructure, analytical skills and resources for data management should be carried out in an equitable manner. Some jurisdictions/sectors making significant contributions towards the evidence base for pandemic response may not have access to state-of-the art information technology and resources. The minimum required infrastructure for pandemic response in terms of technology, interoperability, skills, people and frameworks should be accessible to all jurisdictions/sectors.

2.2.3 FAIR and Timely

The consensus in this series of guidelines and recommendations 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. The FAIR principles (Wilkinson et al., 2016) address a primary concern that has led to the formation of the group writing these guidelines: availability and reusability of research data on COVID-19 in order to prevent unnecessary duplication of work. Many of the specific guidelines in this document address what can be done to make the data as FAIR as possible with a reasonable time investment.

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 perfect 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. FAIRifying data saves valuable time, as it facilitates a level of trustworthiness in research outputs.

Researchers should also be encouraged to share what they have as-is without fear of it being insufficient, and signal that help is needed. The reusability of data can be increased 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 data. This would be a second-phase step that should not unnecessarily slow down researchers collecting data.

In the COVID-19 situation access to data should be as open as possible. This does not necessarily mean completely open access, as data must also be protected as necessary, but measures to control and manage risk (e.g. encryption, anonymisation, de-identification, aggregation, data use agreements) can be used to ease authorised access as much as possible, while still offering adequate protections. 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. Immediate access with licences that are as open as possible is desirable, but effort should be put into the quality and documentation of the dataset.

Finally, it is important to note that a lot of data that are very relevant to the pandemic are kept exclusively on websites and are therefore extremely fragile. This data should be deposited in TDRs, but to address existing gaps, the websites should be web-archived systematically (and permit doing so by way of their robots.txt), so as to ensure persistent availability of the information and to facilitate retrospective analyses. Preference should be given to public web archives that are created and stored by archival organisations.

2.2.4 Data Management Planning

Sharing data in a FAIR and timely way requires planning for data management early in the process of any research undertaking. 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.

Researchers should create a Data Management Plan (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, and it should be updated regularly to ensure data are managed

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

DMPs devise how existing data will be used, how new data will be stored, documented and quality controlled, (and any issues around the handling of sensitive data, legal and ethical issues), and how and where the data will be shared and preserved. DMPs should also identify the human and financial resources required for data management activities.

Researchers should contact, where possible, 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. Working with a dedicated data steward can significantly affect the maturity of this process, and funders and institutions should be encouraged to provide support and recognition for data stewardship roles and contributions.

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 are safeguarded so they can be located, understood, and effectively used and reused. Roles and 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.

2.2.5 Metadata

The key to finding and using digital assets is metadata. Several of the FAIR principles also call for rich 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 generally applicable 'metadata element set' 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), Data Catalog Vocabulary (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. A simple but key recommendation, noted in the COAR recommendations for COVID-19 resources in repositories, is to add the keyword tag "COVID-19" under subject. Metadata for research software is also essential for its reuse and to enable reproducibility.

Providing FAIR access to assets would be much enhanced 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 the future, richer generic metadata standards will be used. For a longer registry of metadata standards, see the Metadata Standards Catalog or the RDA-endorsed FAIRsharing (in the 'Standards' section).

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 section on legal and ethical considerations.

Providers of data sharing infrastructures should perform validation that data complies to recommended metadata/annotation standards in order to help researchers make their data as FAIR as possible.

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).

Critical in the current situation is to have datasets easily findable. Resolvable persistent identifiers like Digital Object Identifiers (DOIs), e.g. linking to a repository or network of repositories, would play a large part in making the data available. Persistent identifiers for primary data sources should be included as a rule in secondary analyses to recognise primary data providers, and this should be requested by publishers and editors.

2.2.6 Documentation

Research outputs need to be well documented, which includes documenting the following: research context, methodologies used to define, construct, and compile data, data cleaning and quality checks, data imputation, data provenance and so on.

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. Reusability of data requires documented provenance: when sharing any secondary data, the generation of which involves comparison against other resources, both the public availability of these used resources and unambiguous referencing of the used resources, including version numbers, should be ensured. It is also useful to document the computing time and resources required for data processing. This could help other researchers to assess the resources required for the computation and help them to decide whether it is feasible to proceed with the local resources available.

Software should provide documentation that describes at least the libraries, algorithms, assumptions and parameters used and software licences or other terms of use.

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 ([International Council on Archives et al., 2020](#)).

2.2.7 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 the long-term preservation and sustained access of their data holdings. Software should be made available in appropriate repositories.

Examples of widely adopted certifications are CoreTrustSeal (CoreTrustSeal), nestor Seal for Trustworthy Digital Archives (nestor) and ISO 16363 (PTAB). Repositories certified by CoreTrustSeal, a result of the RDA Repository Audit and Certification DSA–WDS Partnership WG are listed here. The underlying community-based TRUST principles (Lin et al., 2020) should also be considered.

As the first choice, widely used disciplinary repositories are recommended for maximum accessibility and assessability of the data, as well as repositories that are part of research infrastructures (e.g. CESSDA, ELIXIR, and others), as this also ensures maximum cross-border visibility. These are followed by general or institutional repositories. Using existing open repositories is better than starting new resources.

Making data available in existing and certified repositories will increase the FAIRness of the data. Trustworthy data repositories provide key metadata associated with its datasets, optimally utilising a metadata standard that allows for interoperability. They also employ tools such as persistent identifiers for discovering and citing the data, as well as mechanisms for linking data and other research objects. The re3data.org and the RDA-endorsed FAIRsharing registries can be consulted to find an appropriate repository.

Finally, it is important that policymakers, funders and publishers also promote the use of trustworthy data repositories in their national and institutional policies, calls and data availability policies.

2.2.8 Publications / Data Publications

Rapid publication, i.e. via pre-print repositories or before peer review is possible, along with other forms of knowledge sharing and exchange should be encouraged. Similarly, journals should undergo an expedited review process for pandemic related research. There remains of course the need to balance the rapid dissemination of findings with the dissemination of reliable findings. Full reports should be made openly available immediately upon communication of results, e.g. through a press release.

Research funders and policymakers should implement a "data/software first" publication policy by encouraging the publication of data and software articles in "open" peer-reviewed data journals, or mandating and supporting the deposit of data and associated software in a trustworthy data repository in tandem with the publication of articles. Curated datasets and software, peer-reviewed data and software articles should be treated as first-class research outputs equal in value to traditional peer-reviewed articles.

Funders need to make sure that calls for projects clearly state 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". Publishers need to require publishing of the data, software and code underlying a study, in an even more timely manner than usual. Publishers should actively recommend publishing of data in trustworthy domain-specific repositories where findability is better than in generic or institutional repositories.

3. Data Sharing in Clinical Medicine

3.1 Focus and Description

Health care measures and clinical research are at the forefront of combating the COVID-19 pandemic. Promotion of clinical data sharing is of utmost importance because many studies and trials 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 (Sharing and re-use of individual participant data from clinical trials: principles and recommendations, BMJ Open 2017). Clinical data outside clinical trials (e.g. case studies, descriptive cohorts of patients, etc.) may also be of high value and should be reported.

3.2 Scope

The work highlighted in the Clinical section centres on obtaining consent to address future use of data, conducting clinical trials, sharing the different types of clinical information (personal and health data), and ensuring that results are shared and reused in a trustworthy and efficient manner. Ethics, legal and equity related recommendations are fundamental and applicable to clinical data. However, they can be found in the section on legal and ethical considerations and in other parts of this document.

3.3 Policy Recommendations

3.3.1 Trustworthy Sources of Clinical Data

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

1. Measures should be taken in order to organise the transferral of data and trial documents to a suitable and secure data repository to help ensure that the data are properly prepared, available in the longer term, stored securely (with respect to access control, confidentiality, and integrity) and 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 or accreditation (Consultative Committee for Space Data Systems, 2011; CoreTrustSeal Standards and Certification Board, 2019). A wide range of community-based standardised quality criteria, best practices, and principles (e.g. TRUST Principles (Lin et al., 2020)) should also be considered.

3. If analysis environments that allow in situ analysis of datasets are available, but prevent downloads, they should be provided to the end-user researchers in a pandemic situation, without fees if possible.

4. Tools allowing different datasets 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.

6. Sharing of preclinical data would speed-up the validation of new tools, leading to an update of the current preclinical testing standards used by the pharmaceutical industry. However the reproducibility of preclinical data is a matter of concern (Raimondi et al., 2020).

Data Standards

Using relevant data and metadata standards for clinical data will allow and support the consistent access to and reliable exchange of data from COVID-19 clinical research and case reporting.

1. More support is needed for academic researchers to apply the relevant standards (a 'simplified CDISC' for COVID-19 may be useful); this should be a priority for funders and institutions.
2. 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.
3. Multi-centre and/or multi-country studies, including a sample size calculation according to the primary objective, should be recommended to generate sound evidence on COVID-19 treatments. Policymakers and funders should act so that priority is given to such trials for quickly achieving results. Collaborative trials and multi-arm 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.

FAIR Data

Discoverability and metadata are important elements to optimise sharing and accelerate data use.

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, dataset, a summary of results, publication, data management plan) through one portal (e.g. ECRIN: Clinical Research Metadata Repository (European Clinical Research Infrastructure Network).

2. 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 the pressure to rapidly publish and make data available, there may be a greater risk of data not being properly prepared prior to data sharing. Regardless of the pressure, measures to protect privacy and prevent the risk of re-identification is paramount (e.g. specific data use agreements). For public health emergency situations, some legislation (e.g. GDPR Article 9 (Vollmer, 2018)) contains 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 web page with explicit authority.

Informed Consent for Data Sharing

1. Data and clinical trial information should be made available for broad sharing when possible.

2. 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 (Access Now, 2020).

Publications and Other Formats

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 (Birney et al., 2009).

Biological Samples as Data Sources

1. In the context of a pandemic, 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 with the General Data Protection Regulation (GDPR) provisions on biobanking (Staunton et al., 2019).

2. Main principles are delineated in the Access policy of BBMRI-ERIC, the European Research Infrastructure Consortium for Biobanking and Biomolecular Resources (BBMRI-ERIC, 2020).

Rights, Types and Management of Access

In order to expedite the process of data sharing, standardised 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).

3.4 Guidelines

3.4.1 Data and Metadata Standards for Clinical Data

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, those from the Clinical Data Interchange Standards Consortium (CDISC, 2020a) and, especially for exchanging electronic health records (EHR), HL7 FHIR (Fast Healthcare Interoperability Resources) are particularly encouraged to be considered for ensuring data interoperability. Clinical trials, case reports and public health studies should put the CDISC Interim User Guide for COVID-19 (CDISC, 2020b) 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. Standardised 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) (World Health Organization, 2018),

Systematised Nomenclature of Medicine Clinical Terms ([SNOMED CT](#)) and [LOINC](#) (Logical Observation Identifiers Names and Codes). This ensures unambiguous interpretation (by both humans and computer algorithms) of the used terms describing the data and its elements. SNOMED CT and ICD-10 were both extended by specific terms corresponding to COVID-19 and special use codes were developed for LOINC that can be accessed as pre-release terms.

3.4.2 Clinical Trials on COVID-19

Clinical trials are an important research area to discover and make available safe and effective treatments for COVID-19. International, national, and regional networks exist for clinical trials. Specific resources were also made available to guide clinical trials in COVID-19. 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 adequately protecting individual rights. Platforms that point to them in the various national and international institutions should be further developed and administrations should apply them diligently and transparently.

2. Clinical trials in COVID-19 should be registered at or before the time of first patient enrolment and protocols published in order to favour harmonisation of studies, collaboration among centres, as well as to avoid duplication of efforts.

3. 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.

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) ([International Organization for Standardization, 2015](#)) or the Global Alliance for Genomics and Health (GA4GH) Consent Policy [(Global Alliance for Genomics and Health, 2019)](#).

5. Clinical data and clinical trial information should be done using appropriate reporting guidelines [(see EQUATOR Network guidelines](#) and FAIR Sharing Registry).

6. Multi-centre and/or multi-country studies, including a sample size calculation according to the primary objective, should be performed to generate sound evidence on COVID-19 treatments. Collaborative trials and multi-arm studies comparing different interventions are advisable.

7. Protocols should follow standard criteria for data collection, stratification of the randomised 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.

8. 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-authorisation efficacy and safety studies, after exceptional or conditional approval, should be planned with sponsors in order to favour early access of severe patients to promising medicines.

9. Pre-print publishing and other forms of knowledge sharing and exchange are important to accelerate timely circulation of information.

3.4.3 Immunological, Imaging and Healthcare Data

COVID-19 clinical trials and clinical data of patients infected by SARS-CoV-2 represent valuable information for better knowledge and management of this pandemic. The important elements of such data are clinical presentation and evolution, diagnostic and prognostic data including immunological data, virology test results and imaging, especially lung scan in case of respiratory distress.

All values for metadata and assay results should be defined with the use of domain specific controlled vocabularies. These data standards are recommended for the following data types:

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

Minimal information on flow cytometric data should be provided via the MiFlowCyt minimal standard (Lee et al., 2008). Raw data should be provided in the standardised .fcs format (Spidlen et al., 2010). The primary cytometry data in .fcs format is greatly enhanced by the inclusion of interpreted data (e.g. the cell population name, definition and frequency) (Dunn, 2020a).

Cell population names should be the standard name from a curated reference source (e.g. Cell Ontology). Use of standardised 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. Gating information should be provided using Gating-ML (Spidlen et al., 2015).

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 standardised 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 (Dunn, 2020b).

3. **Neutralising Antibody Titer**

Standardised names for viral targets using reference sources (e.g. National Center for Biotechnology Information [NCBI] Taxonomy) is recommended. Description of the neutralising antibody type (e.g. IgM, IgG) and detector enhances interoperability.

4. **Virus Presence and Titer**

Standardised 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 the work of (Persons et al., 2020) should be applied. Digital Imaging and Communications in Medicine (DICOM) (DICOM Standards Committee)— 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 (DICOM Standards Committee, 2011) offers a framework for de-identification of clinical imaging data for use in research studies (Freymann, 2012).

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 (DICOM Standards Committee).

6. Genomics Data and Health-related Data

Sharing genomic and health-related data should follow recommendations modeled after the "Global Alliance for Genomics and Health (GA4GH) Consent Policy" (Global Alliance for Genomics and Health, 2019). 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 can be found in the legal and ethical sub-group section.

4. Data Sharing in Omics Practices

4.1 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.

4.2 Scope

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

Within this scope, recommendations on data that are already frequently associated with biological research on SARS-CoV-2 and COVID-19 are prioritised.

4.3 Policy Recommendations

4.3.1 Researchers Producing Data

The FAIR data principles 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. Considerations for Omics during the COVID-19 pandemic are:

1. 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.
2. 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.
3. 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.

4.3.2 Policymakers & Funders

Due to the high costs involved with high-throughput genomics, few data are available from low- and middle-income countries (LMIC) and from minority ethnic 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.

4.4 Guidelines

4.4.1 Guidelines for Virus Genomics Data

4.4.1.1 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](#). Some specific examples are:

1. We suggest that raw virus sequence data are stored in one of the International Nucleotide Sequence Database Collaboration ([INSDC) archives](#), as each of these is well known and openly accessible for immediate reuse without undue delays:
    1.1. DNA Data Bank of Japan ([DDBJ](#)) ([Ogasawara et al., 2020](#)) Sequence Read Archive (SRA)
    1.2. [ENA](#) (European Nucleotide Archive at EMBL-EBI), for submission documentation see [ENA Documentation](#) (ENA-Docs, 2020)
    1.3. [NCBI SRA](#) 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](#) accessible through [NCBI Virus](#) ([Hatcher et al., 2017](#)), for submission documentation see [Viral sequence submission documentation](#) (NIH-NCBI, 2020)
    2.2. [DDBJ Annotated/Assembled Sequences](#) (DDBJ, 2020)
    2.3. [ENA](#) (EMBL-EBI, 2008)
3. Virus data submitted to [GenBank](#) (NCBI, 2013; [Benson et al., 2013](#); [Clark et al., 2016](#)) and [RefSeq](#) (NCBI, 2013; [Pruitt et al., 2012](#)) will be available for reuse through [NCBI Virus](#) (NCBI, 2013; [Hatcher et al., 2017](#)).
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.

4.4.1.2 Data and Metadata Standards

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

1. We suggest that data are preferentially stored in the following formats, in order to maximise the interoperability with each other and with standard analysis pipelines:
    1.1. Raw sequences: [.fastq](#) ([Cock et al., 2009](#)); optionally compress with gzip
    1.2. Genome contigs: [.fastq](#) ([Cock et al., 2009](#)); if uncertainties of the assembler can be captured, [.fasta](#) ([Pearson et al., 1988](#)) otherwise; optionally compress with gzip
    1.3. De novo aligned sequences: [.afa](#)
    1.4. Gene Structure: [.gtf](#)
    1.5. Gene Features: [.gff](#)
    1.6. Sequences mapped to a genome: [.sam](#) ([Li et al., 2009](#)) or the compressed formats [.bam](#) or [.cram](#) ([Fritz et al., 2011](#)). Please ensure that the used reference sequence is also

publicly available and that the @SQ header is present and unambiguously describes the used reference sequence.

      1.7. Variant calling: .vcf; please ensure that the used reference sequence is also publicly 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

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

3. For a viral sequence derived from a non-Human host, the viral (if known) and host binomial name should be recorded with the sequence. If that viral sequence was derived from a museum specimen, then the host specimen ID (catalogue number) and specimen holding institution should be recorded, preferably via a PID.

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

      4.1. Newick (Felsenstein, 1986)

      4.2. NEXUS (Maddison et al., 1997)

      4.3. PhyloXML (Han et al., 2009; Stoltzfus et al., 2012)

      4.4. The Minimum Information About a Phylogenetic Analysis (MIAPA) checklist provides a reference list of useful tree annotations (Leebens-Mack et al., 2006; Lapp et al., 2017).

4.4.2 Guidelines for Host Genomics Data

Host genomics data are often coupled to human subjects. This comes with many ethical and legal obligations that are documented in the section on Legal and Ethical Considerations and not repeated here. The COVID-19 host genetics initiative is a bottom-up collaborative effort to generate, share and analyse data to learn the genetic determinants of COVID-19 susceptibility, severity and outcomes.

4.4.2.1 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 and genotyping 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 enable reuse and reproducibility.

4. Researchers interested in human leukocyte antigen (HLA) genomics are referred to the HLA COVID-19 consortium.

4.4.2.2 Repositories[1]

---

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

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 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) synchronise most of their datasets daily[2].
   1.1. Transcriptomics of human subjects (requiring authorised access):
      1.1.1. Database of Genotypes and Phenotypes (dbGaP) (Mailman et al., 2007)
      1.1.2. European Genome-Phenome Archive (EGA) (Lappalainen et al., 2015); the corresponding non-sensitive metadata will be available through EBI ArrayExpress (Athar et al., 2019)
      1.1.3. Japanese Genotype-phenotype Archive (JGA) (Kodama et al., 2015)
   1.2. Transcriptomics (from cell lines/animals):
      1.2.1. ArrayExpress (Athar et al., 2019)
      1.2.2. Gene Expression Omnibus (Barrett et al., 2013)
      1.2.3. Genomic Expression Archive
   1.3. Underlying reads can be retrieved from/will automatically be deposited to the corresponding read archive:
      1.3.1. DDBJ Sequence Read Archive (DRA) (Kodama et al., 2012), for submission documentation see here
      1.3.2. European Nucleotide Archive for submission documentation see here
      1.3.3. NCBI Sequence Read Archive (SRA) for submission documentation see here
   1.4. Microarray-based gene expression data:
      1.4.1. ArrayExpress (Athar et al., 2019)
      1.4.2. Gene Expression Omnibus (Barrett et al., 2013)
      1.4.3. Genomic Expression Archive
   1.5. Data on the originating sample can be retrieved from/will automatically be deposited to the corresponding sample archive:
      1.5.1. DDBJ BioSample
      1.5.2. EBI BioSamples
      1.5.3. NCBI BioSample
   1.6. For specialised 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 specialised resources in addition to one of the resources mentioned above.
2. **Genome-Wide Association Studies** (GWAS):
   2.1. GWAS Catalog
   2.2. EGA (Lappalainen et al., 2015)
   2.3. GWAS Central

---

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

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 et al., 2018](#)) or [VDJServer](#) ([Christley et al., 2018](#)).
    3.2. INSDC repositories via NCBI SRA/Genbank, following the AIRR Community [recommended NCBI submission processes](#).

4.4.2.3 Data and Metadata Standards

1. **Gene Expression Data**
    1.1. Transcriptomics
        1.1.1. Preferred minimal metadata standard [MINSEQE](#)
        1.1.2. Preferred file formats (sequencing-based):
            1.1.2.1. Raw sequences: [.fastq](#) ([Cock et al., 2010](#)), optional compression with gzip or bzip2
            1.1.2.2. Mapped sequences: [.sam](#) compression with [.bam](#) or [.cram](#) ([Fritz et al., 2011](#))
            1.1.2.3. 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 et al., 2001](#))
        1.2.2. Preferred file formats: tab-delimited text, e.g. [MAGE-TAB](#) and [ISA-TAB](#) and raw data file formats from commercial microarray platforms ([Annotare accepted formats](#); [Athar et al., 2019](#))
2. **Genome-wide association studies** (GWAS):
    2.1. Preferred minimal metadata standard: [MIxS](#) ([Yilmaz et al., 2011](#))
    2.2. Preferred file formats:
        2.2.1. Binary files: [.bim, .fam](#) and [.bed](#) ([Chang et al., 2015](#))
        2.2.2. Text-format files: [.ped](#) and [.map](#) ([Chang et al., 2015](#))
3. **Adaptive Immune Receptor Repertoire sequencing** (AIRR-seq):
    3.1. Preferred minimal metadata standards: [MiAIRR](#) ([Rubelt et al., 2017](#))
    3.2. Preferred file formats:
        3.2.1. [AIRR repertoire metadata](#), formatted as .json or .yaml ([Vander Heiden et al., 2018](#))
        3.2.2. [AIRR rearrangements](#), formatted as .tsv ([Vander Heiden et al., 2018](#))

4.4.3 Guidelines for Structural Data

4.4.3.1 Repositories

---

[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 these data are fundamentally distinct from conventional genome sequencing.

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 et al., 2019); a collaborating cluster of three regional centres at (i) for Europe: EBI PDBe (PDBe-KB consortium, 2020) and the Electron Microscopy Data Bank EMDB (Lawson et al., 2011), (ii) for the USA: RCSB PDB (Berman et al., 2000) and (iii) for Japan: PDBj (Kinjo et al., 2017). 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 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 artificial intelligence (AI) and big data
    2.4. Molecular chemical modelling and simulation data using computational tools.

4.4.3.2 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 COVID-19 pandemic is maintained by The Molecular Sciences Software Institute (MolSSI) and BioExcel.

4.4.3.3 Data and Metadata Standards

1. X-ray diffraction
    1.1. There are no widely accepted standards for X-ray raw diffraction data files. Generally these are stored and archived in the vendor's native formats. Metadata are stored in CBF/imgCIF format (in Catalogue of Metadata Resources for Crystallographic Applications).
    1.2. Processed structural information is submitted to structural databases in the PDBx/mmCIF format (Fitzgerald et al., 2006).
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) are deposited and processed through the wwPDB OneDep system (wwPDB Consortium, 2020), following the same annotation and validation workflow also used for X-ray crystallography and nuclear magnetic resonance (NMR) structures. EMDB holds all workflow metadata while PDB holds a subset of the metadata.
    2.3. Most electron microscopy data are 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 et al., 2006).

2.5. Experimental metadata are described in EMDR, see also (Lawson 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 standardisation 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 (Pupier et al., 2018).

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

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).

3.5. Processed structural information is submitted in the PDBx/mmCIF format (Fitzgerald et al., 2006).

4. Neutron scattering

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

4.2. Neutron scattering data are stored in the internationally-adopted ENDF-6 format (Brown et al., 2018) maintained by CSEWG.

4.3. Processed structural information is submitted in the PDBx/mmCIF format (Fitzgerald et al., 2006).

5. Molecular Dynamics (MD) simulations

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 (Goni et al., 2013).

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

6. Computer-aided drug design data

6.1. Virtual screening results are stored in 3D chemical data formats, such as .pdb (Bernstein et al., 1977).

6.2. Structural formulas either in SMILES (Anderson et al., 1987) or IUPAC International Chemical Identifier (InChI), and identified through InChIKey, a non-proprietary identifier for chemical substances (Heller et al., 2015).

4.4.4 Guidelines for Proteomics

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

4.4.4.1 Repositories

1. For a curated list of relevant repositories see FAIRsharing using the query 'proteomics'. The ProteomeXchange Consortium enables searches across the following deposition databases, following common standards.
   1.1. For shotgun proteomics one of:
      1.1.1. PRIDE (Perez-Riverol et al., 2019)
      1.1.2. MassIVE (Wang et al., 2018)
      1.1.3. jPOST (Japan Proteome Standard Repository) (Okuda et al., 2017)
      1.1.4. iProX (integrated Proteome resources) (Ma et al., 2019)
   1.2. For targeted proteomics one of:
      1.2.1. PASSEL (Farrah et al., 2012; Kusebauch et al., 2014)
      1.2.2. Panorama (Sharma et al., 2018; Sharma et al., 2014)
   1.3. For repossessed results one of:
      1.3.1. PeptideAtlas (Deutsch et al., 2009)
      1.3.2. MassIVE (Wang et al., 2018)
2. For recommendations regarding non-mass spectrometry based protein-oriented data (e.g. ELISA, neutralising antibody titers, flow/mass cytometry) see the respective sub-section of the Clinical WG.

4.4.4.2 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 by the HUPO Proteomics Standards Initiative (HUPO PSI) (Taylor et al., 2007; HUPO PSI, 2007) and these are filled using the controlled vocabularies specified by the Proteomics Standards Initiative, PSI CVs.
   1.2. Recommended formats are:
      1.2.1. For gel electrophoresis: gelML (HUPO PSI, 2010)
      1.2.2. For transition lists: TraML (HUPO PSI, 2013)
      1.2.3. For raw spectrometer output: mzML (HUPO PSI, 2017)
      1.2.4. For reporting: mzTab (HUPO PSI, 2014)
      1.2.5. For protein quantisation data: mzQuantML (HUPO PSI, 2017)
      1.2.6. For protein identification data: mzIdentML (HUPO PSI, 2017)
      1.2.7. For metadata ISA-TAB with conversion to PRIDE format

4.4.5 Guidelines 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.

4.4.5.1 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 et al., 2020)
   2.2. Metabolomics Workbench (in the USA) (Sud et al., 2016)
   2.3. Massbank (in Japan) (Horai et al., 2010)

4.4.5.2 Data and Metadata Standards

1. For a curated list of relevant standards see [FAIRsharing using the query 'metabolomics'](). Specific examples:
    1.1. Core Information for Metabolomics Reporting, [CIMR]() standard
    1.2. For identifying chemical compounds use SMILES ([Anderson et al., 1987]()) or InChI ([Heller et al., 2015]())
    1.3. To document Investigation/Study/Assay data, use the [ISA Abstract Model](), also implemented as a tabular format, [ISA-TAB]() in [MetaboLights]() ([Haug et al., 2020]()) and in the [Metabolomics Workbench]() ([Sud et al., 2016]()). For an introduction to ISA, see ([Sansone S-A et al., 2012]()).
    1.4. Recommended formats are:
        1.4.1. For LC-MS data use: ANDI-MS specification ([ASTM International, 2014]()), an analytical data interchange protocol for chromatographic data representation and/or [mzML]() ([HUPO PSI, 2017]())
        1.4.2. For NMR data: [nmrCV](), [nmrML]() ([PhenoMeNal H2020 Project, 2019]())

4.4.6 Guidelines for Lipidomics

Lipidomics revealed an altered lipid composition in infected cells and serum lipid levels in patients with pre-existing 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 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.

4.4.6.1 Generic Recommendations for Researchers

Lipidomics analysis should follow the guidelines of the [Lipidomic Standards Initiative]().

4.4.6.2 Repositories

The recommended repository for lipidomics data is [MetaboLights]() ([Haug et al., 2020]()).

4.4.6.3 Data and Metadata Standards

1. Metadata should follow recommendations from the [CIMR standard]() by the Metabolomics Standards Initiative. It should be made available as tab or comma separated files (.tsv or .csv).
2. Data standards: Data can be stored in LC-MS file, in tab (.tsv) or comma (.csv) separated 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: 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 software such as [LipidBlast](#), [MSPepSearch](#) or [MS-DIAL](#). Finally, lipids classification and nomenclature should follow the [LIPID MAPS guidelines](#).

5. Data Sharing in Epidemiology

5.1 Focus and Description

An immediate understanding of the COVID-19 disease epidemiology is crucial to slowing infections, minimising deaths, and making informed decisions about when, and to what extent, to impose mitigation measures, and when and how to reopen society.

Despite our need for evidence-based policies and medical decision-making, there is no international standard or coordinated system for collecting, documenting, and disseminating COVID-19 related data and metadata, making their use and reuse for timely epidemiological analysis challenging due to issues with documentation, interoperability, completeness, methodological heterogeneity, and data quality.

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, researchers, clinicians, and other potential users.

5.2 Scope

Epidemiology underpins COVID-19 response strategies and public health measures. The recommendations and guidelines support development of an internationally harmonised specification to enable rapid reporting and integration of epidemiology and related data across domains and between jurisdictions.

The guidelines outline a data driven, coordinated global system that encompasses preparedness, early detection, and rapid response to newly emergent threats such as SARS-CoV-2 virus and the COVID-19 disease that it causes.

5.2.1 Supporting Output

The supporting output ([doi.org/10.15497/rda00049](doi.org/10.15497/rda00049)) provides supplemental resources, and further develops the global data-driven vision described in the guidelines. This includes a proposed computable framework to support system responses for emerging pathogens. It offers compatible and reliable data models, protocols, and action plans for newly identified threats such as COVID-19.

5.3 Policy Recommendations

5.3.1 Information Technology and Data Management

Properly funded state of the art infrastructure is required to support advanced research, and to support the data management and data sharing required for rapid response and collaboration (See [Infrastructure Investment](#)). For epidemiology in particular:

1. Ensure an appropriate semantic annotation of data to facilitate its comparability across studies and countries, using as much as possible established standards (e.g. LOINC, UMLS).
2. Rapidly develop standardised tools for aggregating microdata to a harmonised format(s) that can be shared and used while minimising the re-identification risk for individual records.

37

3. Develop machine readable citations and micro-citations for dynamic data. Rapid development of: (a) Resolvable Persistent Identifiers, rather than Uniform Resource Locators (URLs); (b) Machine readable citations; (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).

5.3.2 COVID-19 Epidemiological Data, Analysis and Modelling

1) Implement a global system for early detection and rapid response to emerging zoonoses, integrated across systems for reporting human and animal diseases as well as their vectors (CDC 2020; CGH 2019; eCDC 2019; GLEWS 2006; NASEM 2008, 2009a,b; WHO 2017; WHO, FAO and OIE 2019).
2) Catalogue and document all zoonotic diseases with associated reservoir species and vectors to establish and maintain a global database with potential risks related to humans.
3) Rapidly develop a consensus standard for COVID-19 surveillance data:
   a) Definition of and reporting criteria for COVID-19 testing, reporting on testing, and testing turnaround times.
   b) Policies and definitions: interventions, contact tracing, reporting of cases, deaths, hospitalisations 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, long-term effects in recovered cases, sequelae and immunity, location, demographic, socioeconomic information, and outcome of resolved cases.
   c) Uniform standard daily reporting cut-off time.
4) Rapidly develop an internationally harmonised specification to enable the export/import/integration of epidemiologic data across different levels of data generation (e.g. clinical systems, population-based surveillance/research data, data from biomarker and omics studies, death certification, health insurance data), and successful record-linkage.
5) Develop systems that support workflows to link and share data between different domains, while protecting privacy and security. Use domain specific, time stamped, encrypted person identifiers for this purpose based on industry-standard encryption and cryptographic constructions.
6) Implement internationally harmonised COVID-19 intervention protocols based on peer-reviewed empirical modelling and epidemiological evidence, considering local conditions.
7) Publish situational data, analytical models, scientific findings, and reports used in decision-making and justification of decisions (OGP, 2020b).
8) Account for public health decision-making demands in COVID-19 studies.
9) Harmonise approaches to comparably assess and quantify side-effects of pandemic containment and mitigation measures.
10) Report underlying assumptions and quantify effects of uncertainties on all reported parameters and conclusions for all model predictions etc.
11) Implement a data-driven approach for early identification of hotspots.

5.4 Guidelines

These guidelines highlight current system challenges and offer solutions to help support a larger framework designed to coordinate and structure the collection and use of COVID-19 related data. Six focus areas, described in the guidelines and supporting output (data sources, instruments, privacy, epidemiological data model, computable framework, and an epi-stack architecture), progressively develop a data driven global vision for managing novel biological threats such as COVID-19. We begin with population level data sources that drive the public health strategy and response at all stages of the COVID-

19 threat, from emergence through containment, mitigation, and reopening of society. We then survey clinical and population-based instruments that collect data and discuss preservation of individuals' privacy in shared COVID-19 related data. A full spectrum data model is presented encompassing hospital specific surveillance and electronic health records together with field-based demographic and epidemiological surveillance. We propose Epi-TRACS, a computable framework for emerging pathogen action plans, and an epi-stack that uses the Common Data Model with COVID-19 to integrate clinical data and epidemiological data.

5.4.1 COVID-19 Population Level Data Sources

Although jurisdictions within countries send COVID-19 population level data to the national level, and member countries send data to the WHO, other organisations also collect COVID-19 surveillance data from various sources for a variety of reasons (Table 2). Epidemiologists are thus faced with a situation where it is difficult to assess which datasets are the most up-to-date, complete, and reliable.

See Appendix 1 in supporting output for further details and discussion.

Table 2 - **COVID-19 population level data sources**

SOURCE	DATA
Allen Institute for AI	COVID-19 Open Research Dataset (CORD-19)
Apple Inc	COVID-19-Mobility Trends Reports
European Centre for Disease Control	Geographic distribution of COVID-19 cases worldwide
European Centre for Disease Control	The European Surveillance System (TESSy).
Institute for Health Metrics and Evaluation (IHME)	Global Health Data Exchange (GHDx)
Google Inc.	COVID-19 Community Mobility Report
Johns Hopkins University	COVID19 dataset
Kieren Healy Rpackage	Rpackage - COVID19 Case and Mortality Time Series
University of Oxford	COVID19 dataset
The Atlantic	Tracking Project
The New York Times	Covid-19 Data in the United States
U.N.	Humanitarian Data Exchange (HDX)
U.S. Centre for Disease Control	Cases of COVID19 in the U.S.
University of Washington	Be Outbreak Prepared
World Bank	Understanding the Coronavirus (COVID-19) pandemic through data
World Health Organization (WHO)	Novel Coronavirus (2019-nCoV) situation reports
Worldometer	COVID19 data

5.4.2 Interoperable COVID-19 Epidemiological Surveillance: Clinical and Population-based Instruments

International efforts are currently underway to create COVID-19 instruments/ questionnaires (Tables 3-4). These COVID-specific tools are concentrated at person-level for clinic/hospital surveillance (e.g. Case Report Forms-CRFs), or community surveillance (e.g. questionnaire for general population), and do not necessarily collect the same data. Adherence of new studies to already introduced instruments will strongly enhance the comparability of results.

Table 3 - Questionnaire instruments: Reference studies

COUNTRY	QUESTIONNAIRE
CLINICAL	
Australia	NSW Case questionnaire
Austria	EMS
Europe	TESSy
Germany	Covid-19 research dataset
Uganda	Perinatal COVID-19 Uganda
US	Human Infection with 2019 Novel CoronavirusPerson Under Investigation (PUI) and Case Report Form
Worldwide (WHO member states)	Global COVID-19: clinical platform: novel coronavius (COVID-19): rapid version
POPULATION BASED	
Brazil	Brazil Prevalence of Infection Survey
Europe	Questionnaire by WHO Europe
Germany	GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany
Germany	NAKO COVID-19 Survey tool
Israel	One-minute population wide survey
LMICs	LMIC Covid Questionnaire
South Africa	South African Population Research Infrastructure (SAPRIN) COVID-19 Screening Form
South Asian Countries	National Institute for Health Research (NIHR) Global Health Research Unit
UK	UK COVID-19 Questionnaire

Worldwide (WHO)	[Population-based age-stratified sero-epidemiological investigation protocol for COVID-19 virus infection](#)

Table 4 - Questionnaire instruments: Resources

NIH	[Public Health Emergency and Disaster Research Response (DR2)](#)
NIH	[COVID-19OBSSR Research Tools](#)
PhenX	[PhenX COVID-19 Toolkit](#)

Some of the questionnaire initiatives shown in Tables 3 and 4 are currently feeding into the construction of a COVID-19 demographic and epidemiological surveillance question bank that can be used to form locality specific surveys with both common and distinct questions by domains and cohorts (Wellcome Trust). Some, such as the [UK COVID-19 Questionnaire](#), or the [Covid-19 research dataset](#) are now being funded. Question banks, once they become operational can be queried and filtered by domain, cohort, question text, etc. 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 "localisation" information in the question metadata when questions are reused from one survey to the next.

See Appendix 2 in [supporting output](#) for further details and discussion.

5.4.3 Preservation of Individuals' Privacy in Shared COVID-19 Related Data

Data sharing is essential to improve epidemiological analysis, cross-border pandemic modelling, and coordinated policy development between countries. To ensure privacy, both pseudo-anonymisation of direct identifiers (e.g. patient specific ID's) and anonymisation of indirect identifiers (e.g. socio-demographic information on individuals) must be applied. In addition, it is necessary to control statistical disclosure risk to prevent identification of individuals and their health status using a combination of indirect identifiers such as education level, sex, age, and clinical condition, among others ([Duncan et al., 2011](#); [Templ et al., 2015](#); [Templ, 2017](#)). Using synthetic data may be an option to lower re-identification risks while retaining properties of the original datasets.

See Appendix 3 in [supporting output](#) for further details and discussion.

5.4.4 Full Spectrum View of the COVID-19 Data Domain: An Epidemiological Data Model

The COVID-19 epidemiology that guides public health decisions is dependent on interoperable input data from across a wide variety of domains that include not only clinical, surveillance, research, and modelling data, but also administrative, demographic, socioeconomic, cultural practices and lifestyle, and environmental data, amongst others.

An epidemiological surveillance data model must include 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) personal risk factors.

Standardisation 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 harmonisation.

See Appendix 4 in supporting output for further details and discussion.

5.4.5 Epi-TRACS: Rapid Detection and Whole System Response for Emerging Pathogens

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).

Prior to the COVID-19 outbreak, WHO was already engaged in re-examining GISRS's long-term fitness-for-purpose. In line with these short-term considerations, and with GISRS long-term aspirations, we are recommending a real time, adaptable, rapid response system that supports developing countries, and that employs new technology to combat pandemics and other emerging diseases. The RDA-COVID19-Epidemiology group recommends the creation of a WHO-led EPIdemiological Translational Research Action Coordination System (Epi-TRACS) to add an implementation layer to the existing WHO policies, guidelines, partnerships, and information exchange stack adapted to country-specific contexts.

See Appendix 5 in supporting output for further details and discussion.

5.4.6 COVID-19 Emergency Public Health and Economic Measures Causal Loops: A Computable Framework

Causal loop modeling may be valuable in assessing system sustainability and system resiliency (Bahri 2020; Ricciardi et al. 2020; Wicher 2020). A computable framework in which the actions taken in response to COVID-19 sentinel surveillance can be simulated and assessed both retrospectively and prospectively based on a causal loop diagram may help inform decision making.

See Appendix 6 in supporting output for further details and discussion.

5.4.7 Common Data Models and Full Spectrum Epidemiology: Epi-STACK Architecture for COVID-19 Epidemiology Datasets

Common Data Model (CDM)

Data models may make use of the broad ecosystem of surveillance and clinical data that can also include contact tracing apps, biospecimens, and environmental sample data collected in the community/population or clinic/hospitals.

An emulated trials approach may enable assessment of various risk and prognostic factors (Hernán et.al., 2016). Application of a Common Data Model (CDM) for COVID-19 would facilitate comparing clinical burden and patient outcomes in the context of previous environmental and exposures and comorbidities.

Another possible use case is decision support following an early warning system alert of emergence of a novel pathogen such as SARS-CoV-2. The CDM provides a framework for making public health policy decisions, using partial information about the pandemic that leverages population-level population and health information, person-level epidemiological surveillance information collected in the field and, at the same time or alternatively, person-level patient care information collected in a clinic or hospital setting.

Epi-Stack

The WHO has established the Information Network for Epidemics (Epi-WIN) covering four strategic areas: (a) Identify; (b) Simplify; (c) Amplify; and, (d) Quantify. Evidence is gathered, appraised, and assessed to help form recommendations and policies that have an impact on the health of individuals and population.

The RDA Epi subWG proposes an expanded Epi-Stack feeding into Epi-WIN. This would bring together in a managed system a common data model, the epidemiological surveillance data model, clinical and questionnaire data, population level indicators, and core use cases (Epi-TRACS early warning and response system, decision support research, and patient care research). It is critical that resiliency be built into the system, from the standpoint of how the system functions under stress. When the degree of complexity and interdependencies increase in human made systems, there is always the risk of collapse if not enough balance is built into the system (both the IT infrastructure, data governance, and the "people" part).

See Appendix 7 in supporting output for further details and discussion.

6. Data Sharing in Social Sciences

6.1 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 the 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 sciences research, involving a predominance of observational methods, produces unique data that cannot be recreated in the future. Furthermore, key social sciences data, such as demographics, are valuable tools for all disciplines to be able to understand context and link datasets. 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.

6.2 Scope

Social sciences 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 sciences 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 sciences 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 Behaviours and Behaviour 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 sciences research agenda around COVID-19; (2) promote interoperable cross-disciplinary and cross-cultural data use, collaboration, and understanding; (3) build a foundation for managing social sciences data during pandemics and health emergencies more generally, ensuring that social sciences research can be leveraged for the public good.

6.3 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 plays a crucial role and has value across many stakeholders (e.g. scientists, citizens, governments, private corporations). Data generated from publicly funded projects should be made quickly available to the research community. The following recommendations are aimed at ensuring the policies and practices across the wide array of organisations supporting research during COVID-19 require and ensure high quality, social sciences data in line with FAIR principles.

1. Ensure robust funding streams for social sciences 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 prioritise projects where the social sciences data being produced can be used across domains and are linkable and interoperable.
3. Social sciences funding should require data sharing and provide support for 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. 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.
5. All stakeholders (researchers, research institutions, institutional ethics review boards/ethics committees, healthcare organisations, funding agencies and policymakers) 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.
6. Official statistical agencies and other official data providers 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) and ISO (ISO)).
7. Social sciences journals should require COVID-19 related articles to provide data statements on data availability that point to access in a publicly available repository.

6.4 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 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.

6.4.1 Data Management Responsibilities and Resources

Data management and planning are key, and recommendations Section 2.2.4 should be noted. To ensure broad reuse, a Data Management Plan (DMP) constructed for social sciences data collections should guide the handling of the data over time and help all disciplines (e.g. clinical, epidemiology) understand the data.

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

6.4.2 Documentation, Standards, and Data Quality

Social sciences 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.

Social sciences researchers should be aware of metadata standards used in the social sciences and deposit data into repositories using Data Documentation Initiative (DDI), the Dublin Core Metadata Initiative (DCMI) Scheme, QuDEx, ISO 19115, and SDMX.

Social sciences researchers should be aware of controlled vocabularies and ontologies for the social sciences including Humanities and Social Science Electronic Thesaurus (HASSET), the European Language Social Science Thesaurus (ELSST) which is multilingual, the CESSDA Vocabulary for describing data elements (e.g. analysis unit, data type, mode of collection, etc.), and the DDI set of controlled vocabularies.

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 standardised 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 (American National Standards Institute) and FIPS (Federal Information Processing Standards) for U.S. States and counties, and standard identifiers for organisational entities such as companies (Coffey; INSEAD). This set of actions will facilitate data analysis, harmonisation, linking, visualisation, 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 sciences data it accesses or reuses. For example, the use of readme files to orient a user to a set of files are common in some professions.

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 for social scientists to consult. Two sources for recommended formats are UKDA Recommended Formats (UK Data Service) and Library of Congress Recommended Formats Statement (Library of Congress). The social sciences benefit from the use of many common formats used across disciplines, enabling broader interoperability.

6.4.3 Storage and Backup

Where possible, researchers should avoid using personal storage, and instead use the official storage provisions available from their institution, 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 which offers flexibility and protection, such as a solution offering remote access work (German Data Forum (RatSWD), 2020).

Social sciences data, as is true for human subject 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 licences 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 authorising 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 (See section on Infrastructure Investment).

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

6.4.4 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 reviewing 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 People's 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 licences 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. Licences 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.

When working with commercial partners, seek opportunities to negotiate data sharing mechanisms agreeable to both parties. Develop partner and consortium 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. See [Legal and Ethical Considerations](#) for further details.

6.4.5 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 (CC BY 4.0) or a Creative Commons Public Domain Dedication 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 data repository committed to preservation. As the first choice, disciplinary repositories are recommended for maximum visibility, followed by general or institutional repositories. See Use of Trustworthy Data Repositories and The 5 Safes Model for further guidance on repositories. COVID-19 related social sciences data may be shared in a generalist repository. If you use a general repository (e.g. Figshare Dryad Harvard Dataverse openICPSR Zenodo and others), describe the data using the following as a minimum: the dataset's creator, title, description, year of publication, any embargo, licensing terms, and repository identifier. The COVID Data Repository (ICPSR) accepts data from multiple domains (and formats) as a generalist repository, but because it is run by a social sciences data repository, ICPSR, it offers relevant domain repository benefits (e.g. curation by domain curators, restricted data dissemination options) and ensures social sciences COVID-19 data are FAIR and in a CoreTrustSeal repository.

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. This file should be kept apart from the rest of the data, which provides the linking relationship between any personal identifiers and the randomly assigned unique identifiers.

Social scientists 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.

7. Community Participation and Data Sharing

7.1 Focus and Description

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 mobilise data, add or refine metadata, assess data quality, merge, curate, preserve and combine datasets; analyse, visualise 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.

This section highlights and ultimately supports the work by communities who are collecting, curating and sharing data with the goal of improving research outputs and public knowledge. Employing specific use cases, we detail the achievements and outputs of groups who practise data sharing and stewardship, aiming to broaden access to the existing recommendations and guidelines for research data best practices. As described in "Principles of data sharing in public health emergencies" (GLOPID et al., 2018) and similar publications, such guidelines address issues of data stewardship, ethics and legality in sharing data, technical considerations in making data FAIR, or other similar guidance for collaborating in research during a crisis.

These recommendations and guidelines ultimately aim to facilitate the timely sharing of data relevant to the COVID-19 response, and build much-needed capacity, including knowledge, for similar events in the future. They also hold considerable value for both public and science communication, informing opinions and understanding, whilst supporting decision-making processes.

Although these guidelines have been developed with research data in mind, it is also desirable that data created directly by citizens, patients, communities and other actors in a health emergency be produced, curated and shared in line with the spirit of these stewardship and sharing guidelines. For example, community projects such as OpenStreetMap and Wikidata generate very valuable FAIR and open data (e.g. see Waagmeester et al., 2020), which can be analysed and used along with data from professional research and other sources.

7.2 Scope

This section discusses community participation and is intended to look at data management and sharing issues reflecting on the technical, social, legal and ethical considerations from that perspective.

These recommendations and guidelines are both for and building upon community participation, and the intended audience and contributors are:

1. Researchers undertaking activities along the entire life cycle of pertinent data, especially those not covered in the other RDA COVID-19 WG sections and involving broad-scale community participation but also data stewardship of the community-generated data.
2. Citizen scientists undertaking research activities and in need of guidance (e.g. in terms of ethics) as well as a means to seamlessly contribute to a common body of knowledge and collaborate with other actors involved.

3. Policymakers who are involved in setting the framework for community participation, funding innovation, working on research policy or focusing on integrating data in decision-making.
4. Patients, caregivers and the communities around them that are involved in leveraging data to improve prevention, diagnostics or treatment (this complements the section on Data Sharing in Clinical Medicine).
5. 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).
6. Device makers involved in developing sensors and data generating products for the community to use.
7. Emergency responders, governmental or societal groups involved in prevention and response strategies.
8. 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.
9. Citizens and the public at large, i.e. members of any community - including Indigenous - 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.
10. Other actors (individuals or organisations) who are involved in community-based activities around COVID-19 related data.

This document is intended to provide guidance and recommendations to the groups referenced above, considering their roles and the data challenges they might face:

1. Data subjects: Informed consent - including forms of dynamic consent as necessary - should be obtained from the data subjects before any personal data are 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 controllers: determine the purposes and methods of the processing of personal data, perform the data processing, including analysis, anonymisation and de-identification, storing and preservation, sharing e.g. researchers, app developer, funders, policymakers, health authority.
3. Coordination and information management - documenting decision-making, data workflows. The temporary nature of the disaster response stages that the COVID-19 context asks for often leaves limited time to establish a proper data management workflow and sufficient documentation of the data and decision-making process.
4. Public and science communication - providing easy to access and reliable data and information, dealing with misinformation, develop a proper approach to engage the community participation in data collection.

We anticipate many community participation topics to be relevant for the present COVID-19 context including but 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, digital platforms, apps or other digital tools to enable public participation and/or offer open data. We particularly address two of them as use cases: app development for community-generated data and data challenges in participatory disaster response strategies.

What do we mean by app development for community-generated data?

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

What do we mean by participatory disaster response strategies and the related data challenges?

1. Open government data for disaster response strategies - Recommendations & guidelines for community contributions & engagement, best practices.
2. Citizen scientists and participatory disaster response strategies - Recommendations & guidelines for the engagement of citizen scientists, best practices.
3. Coordination platforms & networks for disaster response strategies.

7.3 Policy Recommendations

7.3.1 Transparency, Community Participation and Data Governance

1. A balance must be achieved between timely testing and contact tracing, exposure notification, emergency response and community safety alongside individual privacy concerns such as surveillance, unauthorised use of personal data and forms of abuse that might result from the identification of subjects.
2. There is a strong need to 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 (Ada Lovelace Institute, 2020).
3. Policymakers need to adopt an active approach to bridging communities and ensuring inputs are streamlined, perspectives from communities are considered, and 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.
4. Preparedness efforts should include provisions to enable web archiving of governmental, public health sector and emergency response websites.

7.3.2 Inclusive, Incremental and Multidisciplinary Approach

1. Consider data and data stewardship expertise as key resources for the detection, investigation and response to public health emergencies such as COVID-19. Encourage and facilitate the participation of data-focused organisations and communities to strategy and response networks.
2. Inclusivity and diversity of roles - ensure developers, data stewards, healthcare professionals, epidemiologists, researchers and the public are represented in the teams driving the development of the data collecting apps, participatory disaster response strategies and coordination platforms. App developers or users or responders with data-related roles 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.

3. 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 Recommendations and Guidelines provide further details on some of these.

4. Whenever possible, aim to reuse and share applicable recommendations that already exist for and from specific communities and/or types of data. To this end, adopt a standardised approach to identify existing guidance from specialised communities. e.g. the Global outbreak alert and response network (GOARN) and its COVID-19 Knowledge Hub covers Capacity Building and Training, Go.Data, Research, Risk communication and community engagement (RCCE) (Global outbreak alert and response network, 2020).

7.3.3 Legal and Ethical Aspects

1. Ethical considerations have to be made regarding the two-way sharing of information using mobile-tracking apps or similar technologies when managing data related to the identification and prevention of infection. These need to be embedded in the emergency response strategies and measures.

2. According to the humanitarian information management principles, information exchange should be a beneficial two-way process between the affected communities and the humanitarian community, including affected governments (Mackintosh, 2000). Therefore, it is crucial to also give timely feedback to communities during the participatory data collection and decision-making process. This requires additional control on data sharing and access management.

3. 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. Data governance comes with accountability and the need to work with the relevant local, national and international authorities to ensure appropriate support networks are in place and the app coordinates with these authorities in such matters.

Use sensitivity to guide technical choices and considerations, such as a decision to only transmit anonymised codes as a means to alert individuals of exposure, tailored to the specific context as much as possible and with additional consideration given to wording, where applicable.

7.3.4 Software Development

Contact tracing apps should adhere to the same development recommendations as other software, particularly to build public trust (see Research Software Sharing for Data Analysis). 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), and the transparency provided by open sharing can also address security concerns.

7.4 Guidelines

In the race against time to collect the data required to combat the COVID-19 pandemic, there is a risk that data are 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, metadata and level of maintenance). 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 principles to ensure data quality, findability, accessibility, interoperability and reuse (FAIR). Implementing the FAIR data principles more widely and in more detail will ease sharing and increase efficiency, which is especially important considering the time constraints we are facing.

7.4.1 Data Collection

1. Encourage public and patient involvement (PPI) throughout the data management lifecycle from the inception of the research question, implementation of the data collection and final data sharing and usage.
2. Ensure apps and participatory response coordination platforms are developed with the research, emergency response and health care questions 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 an 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 the utmost importance when developing applications. Use protocols and methods that aim to protect personal data e.g. Decentralised Privacy-Preserving Proximity Tracing (DP-3T).

7.4.2 Data Quality and Documentation

Follow standardised ways of collecting and curating community-generated data securely and select trustworthy data repositories as a way of standardising COVID-19 data whilst ensuring quality and facilitating sharing.

When collecting and curating the data, ensure detailed metadata are captured with the data and workflows are documented. A consolidated effort should be taken to include the following.

1. Protecting personal data is of key 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, visualisation, 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 technological characteristics and their limitations (e.g. efficiency of the underlying app technology, e.g. Bluetooth versus GPS).
7. Develop, implement and share clear and working protocols/workflows for managing the data processing, especially during participatory crisis response, preferably automatic workflows with

machine actionable data formats to maximise the data processing. Considering the diverse types and sources of data, structured processes for handling data intake and output allow all involved stakeholders to work efficiently.

7.4.3 Data Storage and Long-term Preservation

Consideration for long term storage and preservation of data generated from open government data for disaster response strategies, apps, other resources and response coordination platforms in relation to COVID-19, is not always apparent. For example, what are the retention periods that apply for COVID-19 related data of a specific kind in given legislation? Due to the unprecedented nature of this pandemic, much of this is only being considered on short time frames that do not allow for appropriate planning. In addition to community endorsed best practices in data storage and long-term preservation (see also Section 6.4.3), attention should be focused on:

1. Ensure that prevailing local, national and international legal and ethical requirements for health data and medical studies and open government data, where applicable (e.g. for data retention periods), are adhered to as best as possible.
2. Ensure that provision is made to facilitate 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 retrospective modelling of the current pandemic or in modelling future ones. See 2.2.7 Use of Trustworthy Data Repositories.
4. Data should be available under an open licence that enables reuse, with CC0 as default, unless there are legal and ethical considerations indicating otherwise.
5. Consider the benefits and challenges of either a centralised or decentralised model for data storage and processing.

8. Indigenous Populations and Data Sharing

8.1 Focus and Description

Indigenous Peoples around the globe have diverse narratives of resilience and adaptability; however, they are also acutely impacted by the negative social, economic, environmental and health outcomes of COVID-19 (UN Special Rapporteur on the rights of Indigenous Peoples, 2020). As such, it is vital that Indigenous Peoples are included in all aspects of pandemic-related surveillance, research, research planning, and policy. Systemic policies, and historic and ongoing marginalisation, have led to Indigenous Peoples' mistrust of agencies and the data/research they produce. For example, Indigenous nation-specific COVID-19 data in the United States have been released by government entities without tribal permission and knowledge. These sensitive data continue to be accessed and reused without consent from Indigenous governing bodies by the media, researchers, non-governmental organisations, and others. Although this type of data usage is attempting to combat data invisibility of American Indians and Alaska Natives to address gaps, reporting of tribal-specific data is making tribes more visible in ways that can result in unintentional harm and ignores inherent Indigenous sovereign rights. Media perpetuation of mis-information and dis-information is amplifying confusion and harm to Indigenous Peoples.

To avoid increased distrust and harm, and to improve the quality and responsiveness of data activities, Indigenous data rights, priorities, and interests must be recognised in all COVID-19 research activities throughout the data lifecycle, and in ownership of any resulting innovations. We must also acknowledge that expressions of self-determination vary substantially across nation states due to conditions that also undermine the ability of Indigenous Peoples to govern data or enact sovereignty over data.

The Indigenous Data Guidelines within this document have emerged through global collaborations with Indigenous Peoples and Indigenous data governance advocates. They outline obligations for funders, governments, researchers, and data stewards in the collection, ownership, application, sharing, and dissemination of Indigenous data, specifically in relation to COVID-19 related issues. These Guidelines reflect and support Indigenous Data Sovereignty (see www.GIDA-global.org) and are underpinned by the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP). They do not supersede or replace existing Indigenous governance protocols or agreements developed (or under development) by Indigenous Peoples or nations. Rather, these Guidelines point to the need for Indigenous Peoples and nations to be engaged in governance *on their own terms* across COVID-19 data lifecycles and ecosystems. This demands proactive investment in Indigenous community-controlled data infrastructures to support communality capacity and resilience, and improve the flow of information for effective public health response.

The Indigenous Data Guidelines set out the minimum requirements for Indigenous-designed data approaches and standards, inclusive of Indigenous rights to data governance and decision-making within the planning and design of Indigenous data collection and sharing. The Indigenous Data Guidelines also highlight the inadequacy of personal and individual consent and data privacy protections. For Indigenous Peoples, collective consent and data privacy protections, supported via community-controlled data infrastructure, are essential to ethical Indigenous data practices.

These Indigenous Data Guidelines apply across all sections of the RDA COVID 19 Guidelines and Recommendations.

8.2 Scope

The CARE Principles for Indigenous Data Governance -- Collective benefit, Authority to control, Responsibility, Ethics -- (www.gida-global.org/care) set forth critical considerations for Indigenous rights and interests in data. Indigenous data, in general, comprise data, knowledge, and information that relate to Indigenous Peoples at both the individual and collective level, including data about lands and environment, people, and cultures. In the context of COVID-19, Indigenous data include data about COVID-19 testing (individual and community, e.g. wastewater), cases, hospitalisations, health service access, deaths, and comorbidities, as well as related Indigenous Knowledges about COVID-19, and data on the socioeconomic and environmental correlates and impacts of COVID-19. The CARE Principles provide a framework for the collection, storage, access, and use of Indigenous Peoples' data during the COVID-19 pandemic and beyond.

Access to good quality data is a key driver for the implementation of the FAIR principles – Findable, Accessible, Interoperable, Reusable ([Wilkinson et al., 2016](#)). The FAIR principles are data-centric, supporting greater data findability, accessibility, interoperability and reusability. The FAIR principles facilitate increased data sharing among entities. However, they ignore relationships, power differentials and the historical conditions associated with the collection of data that impact ethical and socially responsible data use. The CARE Principles for Indigenous Data Governance speak to how data are used in ways that are purposeful and oriented towards enhancing the wellbeing of Indigenous Peoples. The CARE Principles can find expression alongside the FAIR principles across data lifecycles from collection to curation, from access to application.

8.3 Policy Recommendations and Guidelines

The CARE Principles for Indigenous Data Governance set a minimum standard for non-Indigenous policymakers, data stewards, researchers, aid groups, and others.

COLLECTIVE BENEFIT: "Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data."

1. **"For inclusive development and innovation"**
   Systemically, existing data ecosystems do not support meaningful inclusion of Indigenous data rights and interests, and when engaged Indigenous Peoples' input is often left out of decision-making, particularly when making data open ([Rainie et al. 2019](#)). Early conscious inclusion at all stages of data lifecycles (design, collection, access, analysis, reporting, storage, protection, use, and reuse of Indigenous data) and throughout data ecosystems (digital infrastructures, analytics, and applications) enhance benefits for Indigenous Peoples and minimize harms such as mis-representation and dis-information.

2. **"For improved governance and citizen engagement"**
   In many countries, Indigenous Peoples are exposed to a higher risk of pandemic-related harm, both to their health and livelihoods. COVID-19 is impacting all communities and responses must recognise the importance of diverse knowledge systems in decision-making in order to advance culturally-informed pandemic policy planning and implementation. By involving Indigenous Peoples throughout the COVID-19 pandemic preparedness and response processes, there is an opportunity to limit negative outcomes and inform both current and future pandemic response planning.

3. **"For equitable outcomes"**

Repositories that include data that are collected or used as part of COVID-19 analyses or responses must explicitly support Indigenous governance of Indigenous data and include provenance for all Indigenous data. All surveillance, research, and data should contribute to addressing Indigenous Peoples' concerns and questions to improve current and future responses, and to achieve equity.

AUTHORITY TO CONTROL: "Indigenous Peoples' sovereign rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and Nations, through their established governing bodies and mechanisms, to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data."

1. **"Recognising rights and interests"**
   Upholding Indigenous rights and interests demands recognition and engagement of Indigenous systems of government and decision-making. Indigenous Peoples' and nations' governing bodies must be formally engaged with prior to the development and implementation of policies and agreements pertaining to Indigenous data that clearly state if, how, and when Indigenous data are collected, analysed, accessed, used/reused, and reported. Permission to use and report on Indigenous Peoples and nations must be granted by appropriate and authorised Indigenous governing bodies. Disclosure of Indigenous information without permission is a violation of Indigenous sovereign rights and undermines Indigenous governance over matters that directly impact Indigenous Peoples.

2. **"Data for governance"**
   Indigenous leadership concerning data collection, ownership, storage, sharing, and use is the defining concept of Indigenous data sovereignty ([Kukutai and Taylor 2016](#)). Indigenous Peoples are in the best position to assess their own needs, priorities, and strengths and are informed by Indigenous responses to COVID-19 (see, for example, [Māori Response Action Plan](#), also see [AIPP COVID-19 Response](#)). As such, Indigenous Peoples need to be supported to lead and/or participate in the design of COVID-19 data systems that involve the collection, analysis, and sharing of Indigenous data. Given that the identification of Indigenous Peoples in data collections has too often led to serious harm and/or stigma, Indigenous Peoples should be able to exercise governance over COVID-19 data that derive from them, individually or collectively, regardless of who collects the data (e.g. government, private sector, researchers), or where they are held. This includes Indigenous data that are de-identified or anonymised for the purpose of sharing.

3. **"Governance of data"**
   Existing Indigenous governance protocols, including those related to decision-making over Indigenous data, must be recognised and adhered to during the COVID-19 pandemic. Indigenous governing bodies must continue to be involved in decision-making on data matters that impact their peoples and nations to ensure collective benefit and minimise harm from Indigenous data.

RESPONSIBILITY: "Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' self-determination and collective benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples."

1. **"For positive relationships"**
   Systemic changes must occur at all levels of government and within institutions that collect, use, or hold Indigenous data to ensure that policies and data sharing agreements are consistent with

Indigenous priorities, are co-determined with Indigenous Peoples, and recognise Indigenous rights to control their data.

2. **"For expanding capability and capacity"**
Indigenous Peoples and nations have often enacted strong, effective first line of responses and defences against COVID-19, proactive investment in Indigenous community-controlled data infrastructure is recommended in order to support community capacity and resilience, and improve the two-way flow of information essential for effective public health responses.

3. **"For Indigenous languages and worldviews"**
Indigenous knowledge and worldviews offer strength for localised contact tracing - local contact tracing data are more likely to be stored in repositories that are governed by Indigenous Peoples. Investments into decentralised contact tracing applications and infrastructure is needed to ensure that Indigenous Peoples can control data as well as narratives over their own contextualised realities.

ETHICS: "Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem."

1. **"For minimising harm and maximising benefit"**
Reporting of identifiable (e.g. ethnic, tribal affiliated, etc.) Indigenous COVID-19 data can contribute to racism and discrimination, hostility, reinforcement of negative stereotypes, and implicitly blame Indigenous Peoples and nations for the spread of COVID-19. Indigenous nations have the responsibility to provide for the safety and welfare of their peoples and nations by determining current and future use of their data, and how and with whom their information will be shared. This is to minimise harm and maximise any benefit that may result from public release of Indigenous-identified COVID-19 data and information. Permission to use and report identifiable Indigenous data by others (e.g. national and state government, researchers, media, etc.) must be granted by Indigenous governing bodies that have the authority to speak on behalf of Indigenous Peoples and nations before their Indigenous data are reported. Disclosure of this information without permission is a violation of Indigenous sovereign rights.

2. **"For justice"**
Indigenous data disaggregation is supported by Indigenous communities (FNIGC, 2016), the United Nations Permanent Forum on Indigenous Issues (2017), and by researchers (Kukutai et al., 2015; Madden et al., 2016). Every effort should be made to collect data that enables Indigenous Peoples to be identified in relation to COVID-19 outcomes *should they desire it*, including the collection of ethnic and tribal identifiers. Non-reporting or aggregation of Indigenous findings into regional populations can disguise the urgent needs of Indigenous Peoples and is insufficient for monitoring the spread of COVID-19 for Indigenous Peoples. Whereas, appropriate reporting and disaggregation is a necessary condition that supports Indigenous visibility and decision-making. However, disaggregated data, without Indigenous governance risks: 1) violation of Indigenous Peoples and nations rights; 2) pejorative judgements from governments, the media, and the public; 3) improper extrapolation of dominant population findings into Indigenous populations; and 4) non-Indigenous algorithms being unreflectively applied to Indigenous data.

3. **"For future use"**
Indigenous data governance is also a prerequisite for determining appropriate future use of data. As contact tracing becomes a key tool to control COVID-19 there has been a noticeable shift from paper-based to electronic tracking, and to increased centralisation. Mobile phone proximity and/or location tracking is another tool being employed by nations and states to mitigate the

spread of COVID-19. While electronic tracking systems may have advantages in their ability to scale and include multiple inputs, they create an enduring record which in many countries do not as yet have an end date. These data, as well as other contact tracing data, can easily be repurposed for other activities. This form of function creep is of particular concern to Indigenous communities who recognise the immediate public health need but face deeper ongoing challenges associated with the use of surveillance as a tool of political oppression. Therefore, Indigenous governance throughout COVID-19 data ecosystems and lifecycles must be supported, including investments in Indigenous community data capacity and infrastructures.

9. Research Software Sharing for Data Analysis

9.1 Focus and Description

It is important to put forward some key practices for the development and (re)use of research software, as doing so facilitates sharing and accelerates the production of results in response to the COVID-19 pandemic.

We provide here a number of foundational, clear and practical recommendations around research software principles and practices, in order to facilitate the open collaborations that can contribute to addressing the current challenging circumstances. These recommendations aim to enable relatively small points of improvement across all aspects of software that will allow its swift (re)use, enabling the accelerated and reproducible research needed during this crisis. These recommendations highlight key points derived from a wide range of work on how to improve the management of software to achieve better research (Akhmerov et al., 2019; Anzt et al., 2020; Clément-Fontaine et al., 2019; Jiménez et al., 2017; Lamprecht et al., 2019; Wilson et al., 2017).

9.2 Scope

These recommendations cover general practices, not details of particular technologies or software development tools. The recommendations in Section 9.5 (Guidelines for Researchers) will not only help researchers improve their software quality and research reproducibility but also have an impact on policymakers, funders and publishers. The aim is that researchers follow the principles as thoroughly as possible, because doing so will improve the research environment for themselves and others. With the recommendations in Section 9.3 (Policy Recommendations), we aim for policymakers and funders to realise the--sometimes behind the scenes--work around research software (e.g. documentation and maintenance). Such awareness will help them to create opportunities addressing, for instance, the acquisition of skills and the full development cycle. With the recommendations in Section 9.4 (Guidelines for Publishers), we aim for publishers to push forward 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 (including, for example, analysis scripts and macros). When necessary to differentiate, we will make an explicit comment.

9.3 Policy Recommendations

Research software is essential for research, and this is increasingly recognised globally by researchers. This section provides recommendations for policymakers and funders on how to support the research software community to respond to COVID-19 challenges, based on existing work (Akhmerov et al., 2019; Anzt et al., 2020). 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.

9.3.1 Support the funding of development and maintenance of critical research software

Policymakers and funders must continue to allocate financial resources to programmes 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, policymakers and funders can increase overall software quality and usefulness. This can be done by making it easier for researchers to move from quick and makeshift coding to creating shared and reusable software, allowing implementation of recommendations detailed in Sections 9.5.4 (provide sufficient metadata/documentation) and 9.5.5 (ensure portability and reproducibility). Funding for software development will also enable anyone producing research software to take the time to produce and document it well, which also aligns with the recommendation in Section 9.5.4. After the software has been delivered, used and recognised by a sufficiently large group of users, human and financial resources should be allocated to support the regular maintenance of the software, for activities such as debugging, continuous improvement, documentation and training.

Examples: UK Research and Innovation is funding COVID-19 related projects that can include work focused 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, 2020c). The Chan Zuckerberg Initiative is funding open source software projects that are essential to biomedical research (Chan Zuckerberg Initiative, 2020).

9.3.2 Encourage research software to be open source and require it to be available

Policymakers should enact policies that encourage software to be available under an open source software licence, or at least require the software to be accessible. All research software that is released under a licence ensures clarity of how it can be used and protects the copyright holders. The use of open source software licences should be seen as the default for research software in publicly funded efforts. If that is the case, 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; it can be modified and redistributed (depending on the licence conditions). 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 testing with different inputs, 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 (Howison et al., 2016) 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 9.5.1 (make your software available), 9.5.2 (release your software under a licence) and 9.5.6 (publish snapshots of your software in an archival repository with persistent identifiers (PIDs)) from the good practices section, thus increasing findability, continued usefulness, and improvement of software.

Examples: The research community has been increasing access to key software and code, with a recent Science article calling for all scientists modelling COVID-19 and its consequences for health and society to rapidly and openly publish their code (Barton et al., 2020). High-profile examples include the Imperial College epidemic simulation model that is being utilised by government decision-makers, and was made publicly available with support by Microsoft to accelerate the process (Adam, 2020). The fact that it was

open also meant it could be inspected and improved. This is an important point that emphasises the raising of quality and the foundation of trust in results.

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

Policymakers and funders should provide programmes 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 and key research support professionals would benefit from the development of core software expertise. Policymakers should support inclusive software skills and training programmes, including development of communities of learners and trainers.

The introduction of such programmes and funding opportunities will increase the overall understanding and adaptation of all recommendations from the good practices section among researchers. This supports the outcomes of the other three recommendations in this section. This also makes it easier for researchers to align to all the recommendations provided in Section 9.5 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 (Carpentries, 2020), USA's NIH events (NIH, 2020d); and the Galaxy Community and ELIXIR's webinar series (ELIXIR, 2020).

9.3.4 Support recognition of the role of software in achieving research outcomes

Policymakers should enact policies and programmes that recognise the important role of research software in achieving research outcomes. It is important that policymakers 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 policymakers work to ensure these systems include proactive responses when these are not implemented. Enacting such policies will encourage researchers to implement recommendations in Sections 9.5.1 (make your software available), 9.5.3 (cite the software you use) and 9.5.6 (publish snapshots of your software in an archival repository with persistent identifiers (PIDs)) from the good practices section, thus creating a self-strengthening system of incentives for the development of high-quality software.

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

9.4 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.

9.4.1 Require that software citations be included in publication

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 scholarly publication 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 (Chue Hong et al., 2017) has been leading work in this area for 3+ years, and currently has a journals task force that is developing sample language for journals to use. The American Astronomical Society (AAS) Journals encourage software citation in several ways (explicit software policy, added the LaTeX \software{} tag to emphasise code used, etc.) (AAS Journals, 2020).

9.4.2 Require that software developed for a publication is deposited in a repository that supports Persistent Identifiers

For publishers to ensure that the research 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 an archival repository that supports PIDs such as Zenodo (CERN, 2020) and Figshare (FigShare, 2020). These repositories provide PIDs that can be directly included in the citation and referenced in a publication, supporting research integrity (Di Cosmo et al., 2018). If the software is deposited along with data (DataCite, 2020), as recommended in certain communities of practice, 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, 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. This is part of the guidance from the FORCE11 Software Citation Implementation Working Group (Chue Hong et al., 2017). The GigaScience journal is another example of publication requiring the availability of software (GigaScience, 2020).

9.4.3 Align submission requirements of software publishers to research software best practices

Recently research software 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, including appropriate software review processes, that encourage and support these practices, for example through adopting or adapting software management statements similarly to the widely adopted data management statements.

Example: The Journal of Open Source Software 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).

9.5 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 response 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 and other researchers in a range of disciplines, writing scripts or macros with almost no background in software development.

9.5.1 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., 2020), Bitbucket (Atlassian, 2020), GitLab (GitLab, 2020), etc.). Furthermore, if using third-party software (proprietary or otherwise), researchers should share and make available the software source code (e.g. analysis scripts) they produce (even if they do not have the intellectual property rights to share the software platform or application itself).

Resources:

Four Simple Recommendations to Encourage Best Practices in Research Software (Jiménez et al., 2017).

FAIR Research Software - code repositories (eScience Center, 2020).

9.5.2 Release your software under a licence

Software is typically protected by Copyright in most countries, with copyright often held by the institution in which the work was performed 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 Licence (Choose a licence, 2020).

9.5.3 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).

9.5.4 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).

9.5.5 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).

9.5.6 Publish snapshots of software in an archival repository with persistent identifiers (PIDs)

Equally important to making the source code available is providing a means of preserving and referring to it in the long-term (Di Cosmo et al., 2018). For this reason, software should be deposited within a repository that supports persistent identifiers (PIDs - a specific example being DOIs), allows for robust metadata and discovery mechanisms, and provides more persistent storage than the code development and collaboration platforms mentioned in Section 8.5.1. Such repositories include Zenodo (CERN, 2020) and Figshare (FigShare, 2020). There are communities of practice that encourage deposition of software (e.g. analysis scripts) and data in one submission. In those circumstances the selected data repository (DataCite, 2020) should provide a PID for the collection. For reproducibility purposes, and if legally allowed, dependencies should also be included in the software deposition. When publishing research results, include a formal citation to the software including a reference to the PID.

10. Legal and Ethical Considerations

10.1 Focus and Description

The intention of these guidelines is to help researchers, practitioners and policymakers deal with the 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 emergencies, it is appropriate 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.

The aim is to identify and collate existing recommendations and guidelines on legal and ethical issues in order to increase the speed of scientific discovery by enabling researchers and practitioners to:

1. Readily identify the guidance and resources they need to support their research work;

2. Understand generic and cross-cutting ethical and legal considerations;

3. Appreciate country- or region-specific differences in policy or legal instruments;

4. Identify the institutional stakeholders best placed to provide relevant ethical and legal guidance.

10.2 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 a 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 ended. 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.

10.3 Policy Recommendations

10.3.1 Initial Recommendations

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 (Omics) 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, ethics review and other formal approval processes should be expedited in compliance with local norms and practices.[4]

6. Policymaking 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, policymakers, 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.

10.3.2 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, 2020a);

---

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

4. The Council of Europe statement on bioethics during COVID-19: ([Council of Europe Bioethics, 2020b](#));

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](#)).

10.4 Guidelines

10.4.1 Cross-Cutting Principles

In addition to following the FAIR principles, all activities, especially in times of pandemic or other public emergencies, should be guided by:

1. The CARE (Collective benefits, Authority to control, Responsibility and Ethics) principles to ensure ethical treatment of individuals and communities ([Global Indigenous Data Alliance, 2019](#));

2. 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](#));

3. The Five Safes of research data governance ([UK Data Service, 2020](#); [Ritchie, 2008](#));

4. Research Integrity guidelines ([ALLEA, 2017](#)).

10.4.2 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 documents: 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 the case of a 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. Intellectual property and associated rights should be taken into account in determining appropriate uses of data and of the innovations derived therefrom.

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 or health-related data that are found in many laws and ethical guidelines include the following:

1. Research projects using human data must be approved by an independent research ethics board (or research ethics committee, or institutional research board) prior to the recruitment of participants and the collection of data, in compliance with local requirements. Research ethics boards have broad powers to approve, reject, require modification of, and terminate research projects.

2. All research projects using human data should comply with local legal obligations as outlined in the following.

3. The obligation to respect confidentiality.

4. The obligation to ensure data accuracy.

5. The obligation to limit the identifiability of personal data as far as possible - including via pseudonymisation techniques.

6. The obligation to use anonymised data instead of personal data, or minimise personal data use, or de-identify where possible.

7. The need to process for a specific, authorised, 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.

8. The obligation to inform individuals about the processing of their data.

9. To hold oneself accountable to, and remain transparent towards, the individuals concerned by the data used.

10. To provide individuals access to their data, and to rectify errors or biases in the data on request.

11. To allow individuals to object to the processing of their data if required by law.

12. 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].

13. The obligation to ensure that data are collected from representative sub-populations and not confined to one group[6].

14. The obligation to ensure equal treatment across cohorts to:

   14.1. Prevent marginalisation of vulnerable groups;

   14.2. Encourage engagement from vulnerable groups;

   14.3. Display trustworthiness and warrant trust ([The South African San Institute, 2017](#)).

15. 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](#)).

16. The obligation to apply legal and ethical practice to all stages of data collection, processing, analysis, reporting and sharing.

17. 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.

18. 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 the identification of specific persons.

19. 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.

20. 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.

10.4.3 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. Researchers attached to an academic institution may find guidance from the following (if available at a particular institution):

   1.1. Research Ethics Boards (REBs), such as an Institutional Review Board (IRB) or Research Ethics Committee (REC), will provide guidance; in some cases, they will review, require modification, approve or stop a research project;

   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 ad hoc mechanisms to support emergency research efforts; or the appropriate governing body (e.g. the National Health Service [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 specialised in research ethics including REBs, and administrators responsible for authorising 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.

10.4.4 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 often 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 need to take into account the possibility of future re-identification, and manage this risk by means of a risk assessment (see Phillips et al., 2016).

In the European Union, for example, 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 (Templ et al., 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 usually fall within the scope of data protection legislation (GDPR, Recital 26) and so therefore require closer controls and management. For the purposes of this document:

1. Anonymised data refers to data where direct and indirect personal identifiers have been removed. Anonymised data poses only a minimal risk of individual re-identification, in considering the context of the data's use and the means reasonably likely to be used to perform re-identification.

2. **De-identified[7] data** refers to data where direct 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 et al., 2016](#)).

3. **Pseudonymised data** refers to 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](#)).

4. **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.

5. **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.

6. **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:

   6.1. **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;

   6.2. **Comorbidities:** for clinical data, where multiple conditions may present for an individual, this also increases the likelihood of being able to identify that individual.

7. **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 et al., 2001](#); [Griffiths et al., 2019](#)).

Our overall **Recommendations on Anonymity include:**

Check with your institution, data protection officer or authority, and institutional review board to determine local definitions of the terms (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.*

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

---

[7] Sometimes referred to as de-personalised

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. In carrying out a re-identification risk 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. The statistical disclosure control methods used, and the re-identification risk assessment, should account for privacy risks to groups and communities. The potential for sensitive attributes to be revealed absent individual re-identification should also be accounted for.

10.4.5 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 specified 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 is 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). Depending on the jurisdiction and the research domain, there may be an additional requirement to seek consent. This may include a representative community board as well as participants themselves.

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.

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:

---

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

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.

10.4.6 Licensing Data and Licensing Software

In releasing data or software for restricted or open use, it is recommended to apply a licence that clarifies the permissions inherent in the data or software. Releasing data or software without an associated licence can create uncertainties as to the permissions inherent in the data that may discourage prospective users from using the data or software. Further details about software licensing can be found in section 9.5.2. For data sharing, it is generally recommended to use an open licence or public dedication to license data that is intended for unrestricted public use.

Choosing the most appropriate licence or similar instrument can be challenging. Certain licences and public domain dedications provide no attribution requirements or use limitations. Examples include CC0 and the Open Data Commons Public Domain Dedication. Other open data licences impose certain limitations on data reuse and can require the attribution of data authorship in a standardised format. Such licences include CC-BY 4.0., the Linux Community Data License Agreement – Sharing, and the Open Data Commons Attribution License. (Bernier et al., 2020). Further documentation to help you choose a data or software licence can be found at www.choosealicence.com (Choose a licence, 2020).

Attribution licences foster accountability on the part of data depositors and can incentivise increased data sharing. Using fully open licences or public domain dedications can promote interoperability and ensure that data will not be subject to incompatible restrictions or use requirements. Data that are anticipated for big data analytic use or use in conjunction with a large number of other datasets, independently sourced, may be best served by a fully open licence that imposes no attribution requirement.

Identifiable personal data and health data can require more restrictive licensing schemes in combination with appropriate data governance to best safeguard the ethico-legal privacy rights of the individuals concerned by the data. Further, it is a recommended best practice to ensure that the licences applied to data are compatible with any contractual or legal obligations of data users, including the obligations imposed by research funding agreements or employment contracts.

10.4.7 The 5 Safes Model

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, along with Statistical Disclosure Control (see below under **Safe Outputs**). Research Data Centres using the 5 Safes model will typically provide different methods of access to data, including remote-only or controlled, on-site connection.

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.

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.

10.4.8 Vulnerable Groups

The overall motivation in producing these guidelines and recommendations has emphasised the open and timely sharing of research data. There is an important consideration, however, when dealing with groups and not just individual participants. Vulnerable groups may include ethnic minorities like Roma or Sinti, or others such as children, migrants or refugees or those with mental or physical disabilities. They often are disproportionately affected by unequal access to health and preventative services. As well as the Indigenous populations discussed above, such groups should be given additional consideration.

First of all, these vulnerable groups should be considered for inclusion in research, clinical trials, testing and epidemiology surveys with the same opportunities as others; individuals in such groups also have the same rights as others to information, access to results where pertinent, and protection of privacy. Specific measures to be inclusive of such groups should be put in place.

This is also true in terms of licensing as well as the collecting, processing and sharing of data (Taylor et al., 2017, Mental Health Europe, 2020). Although a general recommendation would be to use a permissive licence (such as CC 0 mentioned in Section 10.4.6 above), it is important to remember that licences are not aimed at protecting the rights and expectations of individuals or groups represented in the data. For instance, advanced data analytic techniques may identify previously de-identified individuals themselves (Zheng et al., 2011; Bedagkar-Gala et al., 2014), or groupings among individual parties in the dataset which they were unaware of (O'Neil, 2016; see also Boyd et al., 2012). This could lead to stigmatisation and marginalisation (van Aasche et al., 2013). Therefore, when choosing a licence or when reviewing the ethical implications of sharing data, it is important to consider vulnerable groups and ensure their interests are respected. This of necessity includes data which are not typically thought of as personal data. For example, identifying rare vegetation or animals associated with an Indigenous group may help pinpoint their location and therefore expose them to risk.

11. Glossary

This glossary is intended to aid readers in understanding the meaning of selected terms as they are used in this document and does not represent a consensus of CWG on the best definition of each term, nor an attempt to necessarily include all, or even the most important, alternative definitions. The definitions provided here, instead, are intended to reflect the meanings most relevant to the context of this document.

Term	Definition
Access	With regard to research digital resources (i.e. data and software), the continued, available for use, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy and functionality deemed to be essential for the purposes the digital material was created and/or acquired for. Users who have access can retrieve, manipulate, copy, and store copies on a wide range of hard drives and external devices (CASRAI).
Access Controls	Given a data object name, access controls define access relationships between the following metadata: data object name, user name (or user group, or user role), and access permission. The information can be stored as metadata information associated with each data object. The information can be generated dynamically by applying the access controls of the collection that organises the data objects (CASRAI).
Algorithm	In computing, a detailed sequence of steps which, when followed, will accomplish a task (Coltness Computing).
Anonymisation	The process of removing personal identifiers, both direct and indirect, that may lead to an individual being identified. An individual may be directly identified from their name, address, postcode, telephone number, photograph or image, or some other unique personal characteristic. An individual may be indirectly identifiable when certain information is linked together with other sources of information, including, their place of work, job title, salary, their postcode or even the fact that they have a particular diagnosis or condition (University College London).
Anonymised Data	Data where direct and indirect personal identifiers have been removed. Anonymised data poses only a minimal risk of individual re-identification, in considering the context of the data's use and the means reasonably likely to be used to perform re-identification.
Archiving	A curation activity that ensures that data are properly selected, stored, and can be accessed, and for which logical and physical integrity are maintained over time, including security and authenticity. Web archiving follows the same processes to capture web-published data for posterity. (revised from CASRAI).
Assay	To perform an examination on a chemical in order to test how pure it is (Cambridge Dictionary).
Big Data	An evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that have the potential to be mined for information (CASRAI).
Biobank	A repository that stores biological samples and associated information organised in a systematic way for research purposes (revised from ScienceDirect).

Case Report	The scientific documentation of a single clinical observation which describes and analyses the diagnosis and/or the management of one or two patients, with a time-honoured tradition in medicine and scientific publication (revised from NCBI).
Citizen Science	Citizen science is the practice of public participation and collaboration in scientific research to increase scientific knowledge. Through citizen science, people share and contribute to data monitoring and collection programmes (National Geographic).
Clinical Study	Any investigation in relation to humans intended: (a) to discover or verify the clinical, pharmacological or other pharmacodynamic effects of one or more medicinal products; (b) to identify any adverse reactions to one or more medicinal products; or (c) to study the absorption, distribution, metabolism and excretion of one or more medicinal products; with the objective of ascertaining the safety and/or efficacy of those medicinal products (Clinical Trial Regulation N.536/2014)
Clinical Trial	A clinical study which fulfils any of the following conditions: (a) the assignment of the subject to a particular therapeutic strategy is decided in advance and does not fall within normal clinical practice of the Member State concerned; (b) the decision to prescribe the investigational medicinal products is taken together with the decision to include the subject in the clinical study; or (c) diagnostic or monitoring procedures in addition to normal clinical practice are applied to the subjects (Clinical Trial Regulation N.536/2014). WHO designates a clinical trial as follows: 'any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes'.
Cloud Computing	A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualised, dynamically- scalable, managed computing power, storage, platforms and services are delivered on demand to external customers over the internet (CASRAI).
Cohort	A group that is part of a clinical trial or study and is observed over a period of time (National Cancer Institute).
Community Participation	Involves both theory and practice related to the direct involvement of citizens or citizen action groups potentially affected by or interested in a decision or action (Springer Link).
Compassionate Use	A treatment option that allows the use of an unauthorised medicine. Under strict conditions, products in development can be made available to groups of patients who have a disease with no satisfactory authorised therapies and who cannot enter clinical trials (European Medicines Agency).
Compound Library (or: Chemical Library)	A collection of molecules that are synthesised with the aim that they represent a given fraction of the theoretically possible chemical compounds that have yet been made. Research is focused on both the generation of libraries and on new methodology to screen them in the search for new or improved properties (Nature).
Confidential Information	Any information obtained by a person on the understanding that they will not disclose it to others, or obtained in circumstances where it is expected that they will not disclose it (CASRAI).
Confidentiality	The duties and practices of people and organisations to ensure that individual's personal information only flows from one entity to another according to legislated or otherwise broadly accepted norms and policies (CASRAI).

Consent	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. Information concerning the data collection process is presented to the subject or the subject's representative with an opportunity for them to ask questions, after which approval is documented. Consent should be informed (e.g. the individual knows what is going to happen and why) and freely given (without coercion or similar motivation) by someone with capacity, unambiguous and auditable (OECD).
Consent - Clinical	A patient agrees to undergoing a procedure, including taking part in a trial.
Consent - Data Protection	A data subject agrees to personal data being processed for specified purposes.
Consent - Research	A participant agrees to take part in a research study or experiment.
Contact Tracing	The process of monitoring an individual who has been in close contact with a person infected with a virus, who is at higher risk of becoming infected themselves or potentially further infecting others (WHO).
Controlled Vocabulary	A list of standardised terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain (CASRAI).
Copyright	A legal right created by the law of a country that grants the creator of an original work exclusive rights for its use and distribution. International agreements on copyright, such as the Berne Convention, ensure global recognition of national copyrights (CASRAI; Wikipedia).
Cross-cultural	Allows comparison of different cultures.
Cross-morbidities	For clinical data, where multiple conditions may present for an individual, this also increases the likelihood of being able to identify that individual.
Cross-national	Allows comparison of different countries.
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.
Data	Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation (CASRAI).
Data Analysis	A data lifecycle stage that involves the techniques that produce synthesised knowledge from organised information (CASRAI).
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, such as cross-referencing or cross-correlation, or comorbidities.
Data Cleaning	The process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. Cleaning is a continuous process that requires corrective actions throughout the data lifecycle (Sisense; CASRAI).
Data Completeness	The degree to which all required measures are known. Values may be designated as "missing" in order not to have empty cells, or missing values may be replaced with default or interpolated values. In the case of default or interpolated values, these must be flagged as such to distinguish them from actual measurements or observations (CASRAI).

Data Curation	A managed process, throughout the data lifecycle, by which data and data collections are cleansed, documented, standardised, formatted and inter-related. This includes versioning data, or forming a new collection from several data sources, annotating with metadata, adding codes to raw data (CASRAI).
Data Custodian	A data custodian is an IT individual or organisation responsible for the IT infrastructure providing and protecting data in conformance with the policies and practices prescribed by data governance (CASRAI).
Data Element	A unit of data for which the definition, identification, representation (term used to represent it), and permissible values are specified by means of a set of attributes (CASRAI).
Data Imputation	The substitution of estimated values for missing or inconsistent data items (fields). The substituted values are intended to create a data record that does not fail edits (OECD).
Data Integrity	The assurance that information can only be accessed or modified by those authorised to do so. In the context of data quality: The assurance the data are clean, traceable, and fit for purpose. (CASRAI).
Data Linkage (or: Linkage)	The process of combining datasets through code. Two methods are probabilistic and deterministic matching.
Data Management	The activities of data policies, data planning, data element standardisation, information management control, data synchronisation, data sharing, and database development, including practices and projects that acquire, control, protect, deliver and enhance the value of data and information (CASRAI).
Data Management (System) Infrastructure	An infrastructure used to provide data management and enforce data management policies, including resources such as data management plans, a data repository, an information catalogue, devices or hardware, algorithms or software used to store, retrieve and process data (revised from CASRAI).
Data Management Plan (DMP)	A formal statement describing how research data will be managed and documented throughout a research project and the terms regarding the subsequent deposit of the data with a data repository for long-term management and preservation (CASRAI).
Data Mining	The process of analysing multivariate datasets using pattern recognition or other knowledge discovery techniques to identify potentially unknown and potentially meaningful data content, relationships, classification, or trends (CASRAI).
Data Model	A model that specifies the structure or schema of a dataset. The model provides a documented description of the data and thus is an instance of metadata. It is a logical, relational data model showing an organised dataset as a collection of tables with entity, attributes and relations (CASRAI).
Data Preprocessing	Any type of processing performed on raw data to prepare it for another processing procedure. Preprocessing may include: data sampling, data transformation, de-noising, data normalisation, data standardisation, or feature extraction (CASRAI).
Data Preservation (or: Conservation)	An activity within archiving and data management in which specific items of data are maintained over time so that they can still be accessed and understood through changes in technology (CASRAI).
Data Processing	A generic concept referring to all kinds of procedures being executed on data at any point in the data life cycle (CASRAI).

Data Provenance	A type of historical information or metadata about the origin, location or source of the data, or the history of the ownership or location of an object or resource including digital objects (CASRAI).
Data Quality	The reliability and application efficiency of data. It is a perception or an assessment of a dataset's fitness to serve its purpose in a given context (CASRAI).
Data Sharing	The practice of making data available for reuse. This may be done, for example, by depositing the data in a repository, through data publication. (CASRAI).
Data Sharing Agreement (or: Data Transfer Agreement)	An inter-institutional or intra-institutional agreement to share data according to certain terms and conditions. Data sharing agreements identify the parameters which govern the collection, transmission, storage, security, analysis, re-use, archiving, and destruction of data (University of Waterloo).
Data Standard	s The requirements, specifications, guidelines or characteristics that can be used for the description, interoperability, citation, sharing, publication, or preservation of all kinds of digital objects such as datasets, code, algorithms, workflows, software, or papers (FAIRsharing).
Data Steward	Manages and oversees an organisation's data assets to provide data users with high quality data that are easily accessible in a consistent manner. While data governance generally focuses on high-level policies and procedures, data stewardship focuses on tactical coordination and implementation (revised from CASRAI).
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.
Dataset	Any organised collection of data in a computational format, defined by a theme or category that reflects what is being measured/observed/monitored. The presentation of the data in the application is enabled through metadata (CASRAI).
De-identified data	Refers 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 & Knoppers, 2016).
Deposit	The action of uploading a digital copy of a work into a digital repository (CASRAI).
Disciplinary Repository (or: Subject Repository, Domain Repository)	A repository oriented for research output from one or more well defined research domains. All researchers working in certain subject areas can make use of disciplinary repositories – regardless of their affiliation or geographic location (OpenAIRE).
Embargo	In academic publication, the act of submitting data to a repository with the explicit requirement that public data access is delayed.
Encryption	The process of converting data to an unrecognisable or "encrypted" form. It is commonly used to protect sensitive information so that only authorised parties can view it. This includes files and storage devices, as well as data transferred over wireless networks and the internet (TechTerms).
Epidemiology	The study (scientific, systematic, and data-driven) of the distribution (frequency, pattern) and determinants (causes, risk factors) of health-related states and events

	(not just diseases) in specified populations (neighbourhood, school, city, state, country, global). It is also the application of this study to the control of health problems (CDC).
FAIR Principles	A set of guiding principles for scientific data management focused on making data Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016).
General Repository (or: Generalist Repository)	Data repository that is domain agnostic and accepts all data formats.
Genomics	The study of all of a person's genes (the genome), including interactions of those genes with each other and with the person's environment (National Human Genome Research Institute).
Health Disparities	Preventable health differences that are experienced by socially disadvantaged groups.
Indigenous Data	Any facts, knowledge, or information about a Native nation and its tribal citizens, lands, resources, cultures, and communities. Information ranging from demographic profiles, to educational attainment rates, maps of sacred lands, songs, and social media activities. Among others. Indigenous data comprise information and knowledge about our environments, tribal citizens and community members, and our cultures, communities, and interests. The definition encompasses both collective and individual level data (Rainie et al., 2017; Nickerson, 2017).
Indigenous Data Governance	The act of harnessing tribal cultures, values, principles, and mechanisms—Indigenous ways of knowing and doing—and applying them to the management and control of an Indigenous nation's data ecosystem. It is the decision-making and the power to decide how and when Indigenous data are gathered, analysed, accessed and used (Rainie et al. 2017; Walter et al. 2018).
Indigenous Data Sovereignty	The right of Indigenous peoples and tribes to govern the collection, ownership, and application of their own data, to maintain, control, protect and develop their cultural heritage, traditional knowledge and traditional cultural expressions, as well as their right to maintain, control, protect and develop their intellectual property over these (Rainie et al., 2017; Kukutai & Taylor, 2016).
Indigenous Knowledge System	Refers to the systems of understandings, skills and philosophies developed by Indigenous societies that informs broad societal functioning and decision-making. This knowledge is integral to Indigenous life and encompasses language, cultural practices and traditions, systems of classification, resource use practices, social interactions, ritual and spirituality (UNESCO).
Indigenous Peoples	Indigenous communities, peoples and nations are those which, having a historical continuity with pre-invasion and pre-colonial societies that developed on their territories, consider themselves distinct from other sectors of the societies now prevailing on those territories, or parts of them. They form at present non-dominant sectors of society and are determined to preserve, develop and transmit to future generations their ancestral territories, and their ethnic identity, as the basis of their continued existence as peoples, in accordance with their own cultural patterns, social institutions and legal system (Martínez Cobo, 1982).
Informed consent	Consent given by a patient, data subject and/or participant as a result of being told the risks and potential benefits of taking part.

Institutional Repository	A repository affiliated with a specific institution.
Interoperability	The ability of data or tools from non-cooperating resources to integrate or work together with minimal effort (Wilkinson et al., 2016).
Intervention/ treatment	A process or action that is the focus of a clinical study. Interventions include drugs, medical devices, procedures, vaccines, and other products that are either investigational or already available. Interventions can also include non-invasive approaches, such as education or modifying diet and exercise (ClinicalTrials.gov).
Interventional Study (Clinical Trial)	A type of clinical study in which participants are assigned to groups that receive one or more intervention/treatment (or no intervention) so that researchers can evaluate the effects of the interventions on biomedical or health-related outcomes. The assignments are determined by the study's protocol. Participants may receive diagnostic, therapeutic, or other types of interventions (ClinicalTrials.gov).
Legal Interoperability	Occurs among two or more datasets when: the legal use conditions are clearly and readily determinable for each of the datasets, typically through automated means; the legal use conditions imposed on each dataset allow creation and use of combined or derivative products; and users may legally access and use each dataset without seeking authorisation from data rights holders on a case-by-case basis, assuming that the accumulated conditions of use for each and all of the datasets are met (RDA-CODATA, 2016).
Licence	An official document that gives permission to own, do, or use a piece of intellectual property such as a process, product, data, or software. Commonly used licences for open access works include Creative Commons licences. (CASRAI).
Lipidomics	The study of the structure and function of the complete set of lipids (the lipidome) produced in a given cell or organism as well as their interactions with other lipids, proteins and metabolites. (Nature).
Memorandum of Understanding (MOU)	A nonbinding written document that states the responsibilities of each party to an agreement, before the official contract is drafted (LegalDictionary.com).
Metabolomics	The large-scale study of small molecules, commonly known as metabolites, within cells, biofluids, tissues or organisms. Collectively, these small molecules and their interactions within a biological system are known as the metabolome (European Bioinformatics Institute).
Metadata	Literally, "data about data"; data that defines and describes the characteristics of other data, used to improve both business and technical understanding of data and data-related processes (CASRAI).
Metadata Schema	A labelling, tagging or coding system used for recording cataloguing information or structuring descriptive records. A metadata schema establishes and defines data elements and the rules governing the use of data elements to describe a resource (Zhang & Gourley, 2008).
Metadata Semantics	Encompasses controlled vocabularies, taxonomies, thesauri or ontologies and add an interpretive/translational layer (beyond any that might be provided by the syntax), and enable complex hierarchical grouping and querying of the data (FAIRsharing).
Metadata Standard	Metadata schemas adopted by national or international communities, establishing a common way of structuring and understanding data, including principles and

	implementation issues for utilizing the standard. (Clobridge, 2010; University of Pittsburgh).
Metadata Syntax	Defines the representation of information from a conceptual model or schema, and the transmission format, such as XML, CSV or RDF, which facilitate information exchange (FAIRsharing).
Non-Design Data (or: Digital Trace Data)	Refers to data that is collected (often through webscraping) from naturally occurring data such as from social network application.
Observational Study	A type of clinical study in which participants are identified as belonging to study groups and are assessed for biomedical or health outcomes. Participants may receive diagnostic, therapeutic, or other types of interventions, but the investigator does not assign participants to a specific intervention/treatment. A patient registry is a type of observational study (ClinicalTrials.gov).
Omics	High-throughput data from cell and molecular biology.
Ontology	A vocabulary with hierarchies, meaningful relations among concepts, and their constraints that allow classification of data models and data items using the provided terms, concepts, and conceptual structures. Ontologies provide a way of expressing specific domains in a way that enables interoperability based on semantics and logics rather than just formats and agreed metadata (GoFAIR; RDA-CODATA).
Open Access	Making peer reviewed scholarly content freely available via the internet (CASRAI).
Open Access Journal (or: Open Access Publishing)	A journal that makes its articles immediately available online to the reader without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. All the articles in the journal are available open access (CASRAI).
Open Data (or: Free Data)	Data that can be freely used, re-used and redistributed by anyone, subject only, at most, to the requirement to attribute and sharealike (Open Data Handbook).
Open Format (or: Open File Format, Open Data Format)	A format with a freely available published specification which places no restrictions, monetary or otherwise, upon its use, and can be used and implemented by anyone. For example, an open format can be implemented by both proprietary and free and open-source software, using the typical software licences used by each (revised from Open Definition; Wikipedia).
Open Government	A governing culture that holds that the public has the right to access the documents and proceedings of government to allow for greater openness, accountability, and engagement (CASRAI).
Open Science	The practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and method (FOSTER).
Open Source	Referring primarily to software, open source products include permission to use the source code, design documents, or content of the product. Distribution terms of open-source software should allow free redistribution, modifications, and derived works, include the source code, and not be restricted by specific product, software or technology (revised from Open Source Initiative).
Patient Registry	A type of observational study that collects information about patients' medical conditions and/or treatments to better understand how a condition or treatment affects patients in the real world (ClinicalTrials.gov).

Persistent Identifier (PID)	A long-lasting reference to a digital object that gives information about that object regardless what happens to it. Developed to address "link rot," a persistent identifier can be resolved to provide an appropriate representation of an object whether that object changes its online location or goes offline (CASRAI).
Personally Identifiable Information (PII)	Any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information (University of Pittsburgh).
Portability	In computing, the ability of a programme to run on different machine architectures with different operating systems (Coltness Computing).
Preprint Publishing	Preliminary version of an article that has not undergone review but that may be shared for comment. Preprints may be considered as grey literature (CASRAI).
Primary Data	Data that have been created or collected first hand to answer specific research questions (Data Tree).
Privacy	The ability of an individual or group to seclude themselves or information about themselves, and thereby express themselves selectively, the right to be let alone, or freedom from interference or intrusion. Information privacy is the right to have some control over how personal information is collected and used (Wikipedia; IAPP).
Proprietary Format	A file format that a company owns and controls. Data in this format may need proprietary software to be read reliably. Unlike an open format, the description of the format may be confidential or unpublished, and can be changed by the company at any time. Proprietary software usually reads and saves data in its own proprietary format (Open Data Handbook).
Protected Health Information (PHI)	Under the U.S.'s Health Insurance and Portability and Accountability Act (HIPAA), protected health information (PHI) is considered to be individually identifiable information relating to the past, present, or future health status of an individual that is created, collected, or transmitted, or maintained by a HIPAA-covered entity in relation to the provision of healthcare, payment for healthcare services, or use in healthcare operations. This information is often sought out for de-identification in research publication (HIPAA Journal).
Proteomics	Proteomics refers to the study of proteomes, but is also used to describe the techniques used to determine the entire set of proteins of an organism or system, such as protein purification and mass spectrometry (Nature).
Protocol	The written description of a clinical study. It includes the study's objectives, design, and methods. It may also include relevant scientific background and statistical information (ClinicalTrials.gov).
Pseudonymisation	The processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable individual (GDPR, Article 4(3b)).
Pseudonymised Data	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).
Public Health Surveillance	An ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice (WHO).
Quality Control (QC)	The operational techniques and activities used in quality management to fulfil requirements for quality (American Society for Quality).
Raw Data	Data that have not been processed for meaningful use. Although raw data have the potential to become "information," they require selective extraction, organisation, and sometimes analysis and formatting for presentation. As a result of processing, raw data sometimes end up in a database, which enables the data to become accessible for further processing and analysis in a number of different ways (CASRAI).
README File	Along with a repository licence, contribution guidelines, and a code of conduct, helps communicate expectations for and manage contributions to a project, typically including information on what the project does, why it is useful, how users can get started, where users can get help, and who maintains and contributes to the project (GitHub).
Regulatory Body	An organisation appointed by the government to establish national standards for qualifications and to ensure consistent compliance with them (NHS).
Remote Access	Ability for an authorised person to access data on a computer or a network from a geographical distance through a network connection.
Repository	A digital archive collecting and displaying datasets and their metadata. A lot of data repositories also accept publications, and allow linking these publications to the underlying data (OpenAIRE).
Reproducibility (or: Reproducible Research)	Published results that can be replicated using the documented data, code, and methods employed by the author or provider without the need for any additional information or needing to communicate with the author or provider. This can also apply to software and software code (CASRAI).
Research Data	Refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation (European Commission).
Secondary Data	Existing data which are being reused for a purpose other than the one for which it was collected (Data Tree).
Selection Bias	A sample that is not representative of the population.
Self-Determination	All peoples have the rights of self-determination. By virtue of that right, they freely determine their political status and freely pursue their economic, social and cultural development (Charter of the United Nations; International Covenant on Civil and Political Rights; International Covenant on Economic, Social and Cultural Rights, Article 1, para. 1).
Sensitive Data	Data that must be protected against unwanted disclosure. Access to sensitive data should be safeguarded. Protection of sensitive data may be required for legal or ethical reasons, for issues pertaining to personal privacy, or for proprietary considerations (OpenAIRE).
Sensitive Personal Data	Personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade-union membership; genetic data, biometric data

(or: Special Category Personal Data)	processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation (GDPR Article 4(13), (14) and (15), Article 9).
Social Science	Research on human society and social relationships.
Software	A set of instructions, data or programmes used to operate computers and execute specific tasks. Opposite of hardware, which describes the physical aspects of a computer, software is a generic term used to refer to applications, scripts and programmes that run on a device. Software can be thought of as the variable part of a computer, and hardware the invariable part (TechTarget).
Source Code (or: Software Code)	The version of software as it is originally written (i.e., typed into a computer) by a human in plain text (i.e., human readable alphanumeric characters) (Linux Information Project).
Statistical Disclosure Control (SDC)	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).
Structural Biology	The study of the molecular structure and dynamics of biological macromolecules, particularly proteins and nucleic acids, and how alterations in their structures affect their function. Structural biology incorporates the principles of molecular biology, biochemistry and biophysics (Nature).
Structured Data	Data whose elements have been organised into a consistent format and data structure within a defined data model such that the elements can be easily addressed, organised and accessed in various combinations to make better use of the information, such as in a relational database (CASRAI).
Transcriptomics	The study of the transcriptome—the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell—using high-throughput methods, such as microarray analysis (Nature).
Trustworthy Data Repository (TDR) (or: Trusted Data Repository)	A data repository that has been certified, subject to rigorous governance, and committed to longer-term preservation of data holdings.
Version Control (or: Versioning)	System for documenting changes made to files that enable earlier versions to be recalled and referenced.

12. Acronyms

AAS	The African Academy of Sciences
AAS	American Astronomical Society
ACE2	Angiotensin-Converting Enzyme-2
AI	Artificial Intelligence
AIPP	Asia Indigenous Peoples Pact
AIRR	Adaptive Immune Receptor Repertoire
AIRR-seq	Adaptive Immune Receptor Repertoire Sequencing
ALLEA	European Federation of Academies of Science and Humanities
ANDI-MS	The Analytical Data Interchange Format for Mass Spectrometry
ANSI	American National Standards Institute
ASTM	American Society for Testing and Materials
BBMRI-ERIC	Biobanking and Biomolecular Resources Research Infrastructure under the European Research Infrastructure Consortium
BMIC	Trans-National Health Institute BioMedical Informatics Coordinating Committee
BMRB	Biological Nuclear Magnetic Resonance Data Bank
CARE	Collective Benefit, Authority to Control, Responsibility, Ethics (Principles for Indigenous Data Governance)
CBF	Crystallographic Binary File
CBF/imgCIF	Crystallographic Binary File, Crystallographic Information File image
CC-BY	Creative Commons By Attribution license
CC0	Creative Commons Universal Public Domain Dedication
CDC	Centers for Disease Control and Prevention
CDISC	Clinical Data Interchange Standards Consortium
CDM	Common Data Model
CERN	European Organisation for Nuclear Research
CESSDA	Consortium of European Social Sciences Data Archives
ChEBI	Chemical Entities of Biological Interest
CIMR	Core Information for Metabolomics Reporting
COAR	Confederation of Open Access Repositories
CODATA	Committee on Data of the International Science Council
COMEST	World Commission on the Ethics of Scientific Knowledge and Technology
COVID-19	Coronavirus disease
CRF	Case Report Form
CSEWG	Cross-Section Evaluation Working Group
CSV	Comma-Separated Values
CT	Computed Tomography scan
CWG	Research Data Alliance COVID-19 Working Group
CyTOF	Mass Cytometry by Time of Flight
DAA	Data Access Agreement
dbGaP	Database of Genotypes and Phenotypes

DC	Dublin Core
DCAT	Data Catalog Vocabulary
DCC	Digital Curation Centre
DCMI	Dublin Core Metadata Initiative
DDBJ	DNA Data Bank of Japan
DDI	Data Documentation Initiative
DICOM	Digital Imaging and Communications in Medicine
DMP	Data Management Plan
DOI	Digital Object Identifier
DORA	Declaration on Research Assessment
DP-3T	Decentralized Privacy-Preserving Proximity Tracing
DR2	Disaster Research Response
DRA	DNA Data Bank of Japan Sequence Read Archive
DSA-WDS	Data Seal of Approval World Data System
EBI	European Bioinformatics Institute
EBOV	Zaire Ebolavirus
ECDC	European Centre for Disease Prevention and Control
ECRIN	European Clinical Research Infrastructure Network
EDPB	European Data Protection Board
EGA	European Genome-Phenome Archive
EHR	Electronic Health Records
ELISA	Enzyme-Linked Immunosorbent Assay
ELSST	European Language Social Science Thesaurus
EMBL	European Molecular Biology Laboratory
EMBL-EBI	European Bioinformatics Institute of the European Molecular Biology Laboratory
EMDB	Electron Microscopy Data Bank
EMDR	Electron Microscopy Data Resource
ENA	European Nucleotide Archive
ENDF	Evaluated Nuclear Data File
Epi-TRAC	EPIdemiological Transnational Research Action Coalition
Epi-TRACS	EPIdemiological Translational Research Action Coordination System
Epi-WIN	World Health Organization's Information Network for Epidemics
ESIP	Earth Science Information Partnership
EU	European Union
FACS	Fluorescence-Activated Cell Sorting (registered trademark of BD Biosciences)
FAIR	Findable, Accessible, Interoperable, and Reusable
FC	Flow Cytometry (See also: FACS)
FDA	United States Food and Drug Administration
FHIR	Fast Healthcare Interoperability Resources
FID	File Investigator Database format
FID	Free Induction Decay
FIPS	Federal Information Processing Standards

FNIGC	First Nations Information Governance Centre
GA4GH	Global Alliance for Genomics and Health
GDPR	General Data Protection Regulation
gelML	Gel Electrophoresis Markup Language
GESIS	Leibniz Institute for the Social Sciences (Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen)
GHDDI	Global Health Drug Discovery Institute of China
GHDx	Global Health Data Exchange
GIDA	Global Indigenous Data Alliance
GISRS	World Health Organization's Global Influenza Surveillance Response System
GLOPID	Global Research Collaboration for Infectious Disease Preparedness
GO	Gene Ontology annotation
GOARN	Global Outbreak Alert and Response Network
GWAS	Genome-Wide Association Study
HASSET	Humanities and Social Science Electronic Thesaurus
HDX	Humanitarian Data Exchange
HIPC	Human Immunology Project Consortium
HL7	Health Level 7
HL7 FHIR	Health Level 7 Fast Healthcare Interoperability Resources
HLA	Human Leukocyte Antigen
HTML	HyperText Markup Language
HUPO	The Human Proteome Organization
IBC	International Bioethics Committee
ICD	International Classification of Diseases
ICPSR	Inter-university Consortium for Political and Social Research
ICU	Intensive Care Unit
IG	Interest Group
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IHME	Institute for Health Metrics and Evaluation
imgCIF	Crystallographic Information File image
InChI	International Union of Pure and Applied Chemistry International Chemical Identifier
InChIKey	Fixed-length format derived from the International Chemical Identifier
INSDC	International Nucleotide Sequence Database Collaboration
INSEAD	Institut Européen d'Administration des Affaires (European Institute of Business Administration)
iProX	Integrated Proteome Resources
IRB	Institutional Review Board
ISA	Investigation/Study/Assay
ISA-TAB	Investigation Study Assay Tabular
ISARIC	International Severe Acute Respiratory and Emerging Infection Consortium
ISO	International Organization for Standardization

ISO/TS	International Organization for Standardization Technical Specification
IT	Information Technology
IUPAC	International Union of Pure and Applied Chemistry
JEFF	Joint Evaluated Fission and Fusion File
JEFF of OECD/NEA	Joint Evaluated Fission and Fusion File of the Organisation for Economic Co-operation and Development/Nuclear Energy Agency
JGA	Japanese Genotype-phenotype Archive
JOSS	Journal of Open Source Software
jPOST	Japan Proteome Standard Repository/Database
JSON-LD	JavaScript Object Notation for Linked Data
LC-MS	Liquid Chromatography - Mass Spectrometry
LMIC	Low- and Middle-Income Countries
LOINC	Logical Observation Identifiers Names and Codes
MAGE-TAB	MicroArray Gene Expression Tabular format
MassIVE	Mass Spectrometry Interactive Virtual Environment
MBAA	Multiplex Bead Array Assay
MD	Molecular Dynamics
MERS	Middle East Respiratory Syndrome
MiAIRR	Minimal Information about Adaptive Immune Receptor Repertoires
MIAME	Minimum Information About a Microarray Experiment
MIAPA	Minimum Information About a Phylogenetic Analysis
MIAPE	Minimum Information About a Proteomics Experiment
MINSEQE	Minimum Information about a Next-generation Sequencing Experiment
MIxS	Minimum Information about any (x) Sequence
mmCIF	Macromolecular Crystallographic Information File
MolSSI	The Molecular Sciences Software Institute
MOU	Memorandum of Understanding
MRI	Magnetic Resonance Imaging
MS-DIAL	Mass Spectrometry universal program for Untargeted Metabolomics
MSPepSearch	Mass Spectrometry Proteomics and Metabolomics program
mzIdentML	Protein Identification Data Markup Language
mzML	Mass Spectrometer Output Markup Language
mzQuantML	Protein Quantization Data Markup Language
mzTab	Tab-delimited file format for Mass Spectrometer-Derived Omics Data
NCBI	National Center for Biotechnology Information
NCBI SRA	National Center for Biotechnology Information's Sequence Read Archive
NEA	Nuclear Energy Agency
NHS	National Health Service
NIH	National Institutes of Health
NIHR	National Institute for Health Research
NMR	Nuclear Magnetic Resonance

NMR-STAR	Nuclear Magnetic Resonance Spectroscopy Self-defining Text Archival and Retrieval Format
nmrCV	Nuclear Magnetic Resonance Controlled Vocabulary
nmrML	Nuclear Magnetic Resonance Spectroscopy Markup Language
NSW	New South Wales
NUTS	Nomenclature of Territorial Units for Statistics
OECD	Organisation for Economic Co-operation and Development
OGP	Open Government Partnership
OMOP	Observational Medical Outcomes Partnership
OPIDoR	Optimiser le Partage et l'Interopérabilité des Données de la Recherche
PASSEL	PeptideAtlas Selected Reaction Monitoring Experiment Library
PDB	Protein Data Bank
PDBe	Protein Data Bank in Europe
PDBe-KB	Protein Data Bank in Europe - Knowledge Base
PDBj	Protein Data Bank Japan
PGD	Plan de Gestión de Datos
PHI	Protected Health Information
PID	Persistent Identifier
PII	Personally Identifiable Information
PPI	Public and Patient Involvement
PRIDE	Proteomics Identifications Database
PSI	Proteomics Standards Initiative
PSI CV	Proteomics Standards Initiative Controlled Vocabulary
PTAB	Primary Trustworthy Digital Repository Authorisation Body
PUI	Persons Under Investigation
QC	Quality Control
QuDEx	Qualitative Data Exchange Schema
RCCE	Risk Communication and Community Engagement
RCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank
RDA	Research Data Alliance
RDF	Resource Description Framework
REB	Research Ethics Board
REC	Research Ethics Committee
ReFRAME	Repurposing, Focused Rescue, and Accelerated Medchem compound library
SAPRIN	South African Population Research Infrastructure
SARS	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SDC	Statistical Disclosure Control
SDMX	Statistical Data and Metadata eXchange
SER	Serialised/Serial Spectra file
SHARC	Research Data Alliance Sharing Rewards and Credit Interest Group
SMILES	Simplified Molecular-Input Line-Entry System

SNOMED CT	Systematized Nomenclature of Medicine - Clinical Terms
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
SRA	Sequence Read Archive
TDR	Trustworthy Data Repository
TESSy	The European Surveillance System
TPM	Transcripts Per Million
TraML	Targeted Mass Spectrometry Method Markup Language
TRUST	Transparency, Responsibility, User focus, Sustainability, Technology Principles for digital repositories
TSV	Tab-Separated Values
UKDA	United Kingdom Data Archive
UMLS	Unified Medical Language System
UN	United Nations
UNDRIP	United Nations Declaration on the Rights of Indigenous Peoples
UNESCO	The United Nations Educational, Scientific and Cultural Organization
UniProt	Universal Protein Resource
URL	Uniform Resource Locator
US HIPAA	United States Health Insurance Portability and Accountability Act
US or USA	United States of America
VGAS	Viral Genome Annotation System
W4M	Workflow for Metabolomics
WG	Working Group
WHO	World Health Organization
wwPDB	Worldwide Protein Data Bank
XML	eXtensible Markup Language
XSD	eXtensible Markup Language Schema Definition

13. Additional Resources

General resources on Covid-19	
Description of the resource	Link to the resource
COVID-19 Data Portal brings together relevant datasets submitted to EMBL-EBI and other major centres for biomedical data.	http://www.covid19dataportal.org/
OpenAIRE COVID-19 Gateway	https://beta.covid-19.openaire.eu/
LitCovid curated literature hub the 2019 novel Coronavirus providing central access to relevant articles in PubMed	https://www.ncbi.nlm.nih.gov/research/coronavirus/

Resources on data sharing in clinical medicine	
Description of the resource	Link to the resource
Database of publicly and privately funded clinical studies conducted around the world	https://apps.who.int/iris/bitstream/handle/10665/76705/9789241504294_eng.pdf;jsessionid=49F2FD87378AFCA5B22425655E2D0334?sequence=1 https://www.who.int/ictrp/en/ https://clinicaltrials.gov/ https://www.clinicaltrialsregister.eu/ctr-search/search?query=COVID-19 https://www.covid19-trials.com/ https://celltrials.org/public-cells-data/all-covid-19-clinical-trials/79 https://covid19.trialstracker.net/about.html https://www.covid-trials.org/ https://www.coronaclinicaltrials.com/ https://www.cochranelibrary.com/central/about-central https://www.ecrin.org/covid-19-trials-registries https://www.nihr.ac.uk/covid-studies/ https://solidarites-sante.gouv.fr/IMG/pdf/covid19_projets-recherche_therapeutiques.pdf https://www.aifa.gov.it/sperimentazioni-cliniche-covid-19
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

	https://www.egi.eu/egi-call-for-covid-19-research-projects/
Research on "Sharing and reuse of individual participant data from clinical trials: principles and recommendations"	https://bmjopen.bmj.com/content/7/12/e018647 https://vivli.org/
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 guidelines for COVID-19	https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf https://covid19treatmentguidelines.nih.gov/introduction/ https://www.ecdc.europa.eu/en/covid-19-pandemic
Infrastructures and Networks	https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric_en https://www.ecrin.org/ https://www.eu-stands4pm.eu
Regulatory documents	http://www.icmra.info/drupal/ http://www.icmra.info/drupal/sites/default/files/2020-04/Summary%20of%20ICMRA%20meeting_Observational%20studies%20and%20RWE.pdf https://www.fda.gov/emergency-preparedness-and-response/counterterrorism-and-emerging-threats/coronavirus-disease-2019-covid-19 https://www.fda.gov/vaccines-blood-biologics/investigational-new-drug-ind-or-device-exemption-ide-process-cber/recommendations-investigational-covid-19-convalescent-plasma https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19#what's-new-section https://www.ema.europa.eu/en/documents/other/mandate-objectives-rules-procedure-covid-19-ema-pandemic-

	task-force-covid-etf_en.pdf
	https://www.ema.europa.eu/en/documents/other/call-pool-eu-research-resources-large-scale-multi-centre-multi-arm-clinical-trials-against-covid-19_en.pdf
	https://ss.pmda.go.jp/en_all/search.x?q=COVID&ie=UTF-8&page=1
	https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificresearchcovid19_en.pdf
Standardisation of clinical trials and evidence-based approach to clinical research in COVID-19	https://www.spirit-statement.org/
	http://www.comet-initiative.org/Resources
	https://www.cochrane.org/coronavirus-covid-19-cochrane-resources-and-news
	https://www.cebm.net/covid-19/
	https://www.acrrm.org.au/about-us/news-events/news/article/2020/04/06/national-covid-19-clinical-evidence-taskforce-launch-of-clinical-guidelines
	https://covid19evidence.net.au/
Health care and Clinical Data	https://transmartfoundation.org/covid-19-community-project/ https://www.covid19healthsystem.org/mainpage.aspx https://isaric.tghn.org/covid-19-clinical-research-resources/ https://covid19treatmentguidelines.nih.gov/introduction/ https://www.ersnet.org/the-society/news/novel-coronavirus-outbreak--update-and-information-for-healthcare-professionals https://education.aaaai.org/sites/default/files/Suggestions%20or%20Considerations%20for%20Resuming%20Practices.pdf https://www.cdc.gov/coronavirus/2019-ncov/hcp/therapeutic-options.html https://www.ecdc.europa.eu/en/covid-19-pandemic http://www.iedb.org/home_v3.php https://immport.org/ https://immunespace.org/

	https://clarivate.com/cortellis/article/covid-19-testing-fda-guidance-update/?utm_campaign=clarivate& utm_content=Clarivate_Analytics_Organic_Social_Media_S ocial_XBU_Global_2019&utm_medium=Clariv ate&utm_source=clarivatesprout https://apps.who.int/iris/bitstream/handle/10665/331866 /WHO-2019-nCoVSci_BriefImmunity_passport2020.1eng.pdf?sequence =1&isAllowed=y https://apps.who.int/iris/bitstream/handle/10665/50241/ bulletin_1992_70%286%29_699-703.pdf?sequence=1&isAllowed=y https://mrctcenter.org/wp-content/uploads/2020/04/zarin-lau-interpreting-dx-tests-for-COVID.pdf https://www.cdc.gov/coronavirus/2019-ncov/hcp/therapeutic-options.html

Resources on data sharing in omics practices	
Description of the resource	Link to the resource
RDA COVID-19 Omics group	https://www.rd-alliance.org/groups/rda-covid19-omics

Resources on data sharing in epidemiology	
Description of the resource	Link to the resource
RDA COVID-19 Epidemiology group	https://www.rd-alliance.org/groups/rda-covid19-epidemiology
RDA COVID-19 detailed Epidemiology supporting document	https://doi.org/10.15497/rda00049

Resources on data sharing in social sciences	
Description of the resource	Link to the resource
RDA COVID-19 Social Sciences group	https://www.rd-alliance.org/groups/rda-covid19-social-sciences
Best practices for software and trustworthy analysis	https://docs.google.com/document/d/14Cd1cOS8Cv8H hLEkVyPv2UfunvHdLsUaNk7U7AdDJk8/edit?usp=sharin g
Best Practices for measuring the social, behavioural, and economic impact of epidemics	https://deepblue.lib.umich.edu/handle/2027.42/15468 2

Social Psychological Measurements of COVID-19: Coronavirus Perceived Threat, Government Response, Impacts, and Experiences Questionnaires	https://psyarxiv.com/z2x9a/

Resources on community participation and data sharing	
Description of the resource	Link to the resource
RDA COVID-19 Community Participation group	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

Resources on research software and data sharing	
Description of the resource	Link to the resource
RDA COVID-19 Software group	https://www.rd-alliance.org/groups/rda-covid19-software

Resources on legal and ethical compliance	
Description of the resource	Link to the resource
RDA COVID-19 Legal and Ethical group	https://www.rd-alliance.org/groups/rda-covid19-legal-ethical
Global Alliance for Genomics and Health - Responsible Data Sharing to Respond to the COVID-19 Pandemic: Ethical and Legal Considerations	https://docs.google.com/document/d/1wK_NoNYXKy0ttTQ-ySHh3ZRpvPrLV4uPwV8FSq6BQ60/edit

14. 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.

AAS Journals. "Software Citation Suggestions." *AAS Journals* (blog), 2020. https://journals.aas.org/software-citation-suggestions/.

Aasche, Kristof van, Serge Gutwirth, and Sigrid Sterckx. "Protecting Dignitary Interests of Biobank Research Participants: Lessons from Havasupai Tribe v Arizona Board of Regents." *Law, Innovation & Technology* 5, no. 1 (2013): 54–84. https://doi.org/10.5235/17579961.5.1.54.

Abd-Alrazaq, Alaa, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. "Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study." *JOURNAL OF MEDICAL INTERNET RESEARCH* 22, no. 4 (April 21, 2020). https://doi.org/10.2196/19016.

Abdulmajeed, Kabir, Monsuru Adeleke, and Labode Popoola. "Online Forecasting of COVID-19 Cases in Nigeria Using Limited Data." *Data in Brief*, May 2020, 105683. https://doi.org/10.1016/j.dib.2020.105683.

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.

Ada Lovelace Institute. "COVID-19 Rapid Evidence Review: Exit through the App Store?," July 5, 2020. https://www.adalovelaceinstitute.org/our-work/covid-19/covid-19-exit-through-the-app-store/.

Adam, David. "Special Report: The Simulations Driving the World's Response to COVID-19." *Nature* 580, no. 7803 (April 2, 2020): 316–18. https://doi.org/10.1038/d41586-020-01003-6.

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.

Ahn, Matae, Danielle E. Anderson, Qian Zhang, Chee Wah Tan, Beng Lee Lim, Katarina Luko, Ming Wen, et al. "Dampened NLRP3-Mediated Inflammation in Bats and Implications for a Special Viral Reservoir Host." *Nature Microbiology* 4, no. 5 (May 2019): 789–99. https://doi.org/10.1038/s41564-019-0371-3.

AI challenge with AI2, CZI, MSR, Georgetown, NIH, and The White House. "COVID-19 Open Research Dataset Challenge (CORD-19)," 2020. https://kaggle.com/allen-institute-for-ai/CORD-19-research-challenge.

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.

Aitsi-Selmi, Amina, and Virginia Murray. "Protecting the Health and Well-Being of Populations from Disasters: Health and Health Care in The Sendai Framework for Disaster Risk Reduction 2015-2030."

Prehospital and Disaster Medicine 31, no. 1 (2016): 74–78. https://doi.org/10.1017/S1049023X15005531.

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

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.

Allen, T., K.A. Murray, C. Zambrana-Torrelio, S.S. Morse, C. Rondinini, M. Di Marco, N. Breit, K.J. Olival, and P. Daszak. "Global Hotspots and Correlates of Emerging Zoonotic Diseases." *Nature Communications* 8, no. 1 (2017). https://doi.org/10.1038/s41467-017-00923-8.

Allocati, N., A. G. Petrucci, P. Di Giovanni, M. Masulli, C. Di Ilio, and V. De Laurenzi. "Bat–Man Disease Transmission: Zoonotic Pathogens from Wildlife Reservoirs to Human Populations." *Cell Death Discovery* 2, no. 1 (June 27, 2016): 1–8. https://doi.org/10.1038/cddiscovery.2016.48.

Andersen, Kristian G., Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes, and Robert F. Garry. "The Proximal Origin of SARS-CoV-2." *Nature Medicine* 26, no. 4 (April 2020): 450–52. https://doi.org/10.1038/s41591-020-0820-9.

Anderson, Eric, Gilman D. Veith, David. Weininger, and Environmental Research Laboratory. "SMILES, a Line Notation and Computerized Interpreter for Chemical Structures." EPA: Environmental Research Brief. Duluth, MN: U.S. Environmental Protection Agency, Environmental Research Laboratory, 1987. /z-wcorg/. https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=2000CAUR.PDF.

Anderson, R.M., C. Fraser, A.C. Ghani, C.A. Donnelly, S. Riley, N.M. Ferguson, G.M. Leung, T.H. Lam, and A.J. Hedley. "Epidemiology, Transmission Dynamics and Control of SARS: The 2002-2003 Epidemic." *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, no. 1447 (2004): 1091–1105. https://doi.org/10.1098/rstb.2004.1490.

Angelopoulos, Anastasios Nikolas, Reese Pathak, Rohit Varma, and Michael I. Jordan. "On Identifying and Mitigating Bias in the Estimation of the COVID-19 Case Fatality Rate." *Harvard Data Science Review*, 2020. https://hdsr.mitpress.mit.edu/pub/y9vc2u36/release/2.

Anzt, Hartwig, Felix Bach, Stephan Druskat, Frank Löffler, Axel Loewe, Bernhard Y. Renard, Gunnar Seemann, et al. "An Environment for Sustainable Research Software in Germany and beyond: Current State, Open Challenges, and Call for Action." *F1000Research* 9 (April 27, 2020): 295. https://doi.org/10.12688/f1000research.23224.1.

Apple. "COVID-19 - Mobility Trends Reports." AppleMaps, 2020. https://www.apple.com/covid19/mobility.

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.

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

———. "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, 2014. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

Asia Indigenous Peoples Pact. "AIPP's Statement in Solidarity with Indigenous Peoples in Mindanao." *Asia Indigenous Peoples Pact* (blog), May 15, 2020. https://aippnet.org/aipps-statement-in-solidarity-with-indigenous-peoples-in-mindanao/.

———. "COVID-19 and Humanity Lessons Learned from Indigenous Communites in Asia," 2020. https://aippnet.org/wp-content/uploads/2020/04/Combined-2nd-flash-Brief-C19.pdf.

———. "COVID-19 Response." *Asia Indigenous Peoples Pact* (blog), 2020. https://aippnet.org/covid-19-response/.

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.

Atlassian. "Bitbucket | The Git Solution for Professional Teams." Bitbucket, 2020. https://bitbucket.org/product.

Austin, Claire C, Anna Widyastuti, and the RDA-COVID19-WG. "COVID-19 Population Level Data Sources: Review and Analysis." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

Bahri, Muhamad. "The Nexus Impacts of the COVID-19: A Qualitative Perspective," May 3, 2020. https://doi.org/10.20944/preprints202005.0033.v1.

Bai, Zhihua, Yue Gong, Xiaodong Tian, Ying Cao, Wenjun Liu, and Jing Li. "The Rapid Assessment and Early Warning Models for COVID-19." *Virologica Sinica*, April 1, 2020, 1–8. https://doi.org/10.1007/s12250-020-00219-0.

Bajwa, Sukhminder Jit Singh, Rashi Sarna, Chashamjot Bawa, and Lalit Mehdiratta. "Peri-Operative and Critical Care Concerns in Coronavirus Pandemic." *INDIAN JOURNAL OF ANAESTHESIA* 64, no. 4 (April 2020): 267–74. https://doi.org/10.4103/ija.IJA_272_20.

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.

Barton, C. Michael, Marina Alberti, Daniel Ames, Jo-An Atkinson, Jerad Bales, Edmund Burke, Min Chen, et al. "Call for Transparency of COVID-19 Models." Edited by Jennifer Sills. *Science* 368, no. 6490 (May 1, 2020): 482.2-483. https://doi.org/10.1126/science.abb8637.

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-ERIC. "Access Policies | BBMRI-ERIC: Making New Treatments Possible," 2020. https://www.bbmri-eric.eu/services/access-policies/.

———. "BBMRI-ERIC RESPONDS TO THE CORONAVIRUS PANDEMIC: RESOURCES FROM BIOBANKS ACROSS EUROPE AVAILABLE FOR RESEARCH ON COVID-19," April 22, 2020. https://www.bbmri-eric.eu/covid-19.

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 Resource," 2020. FAIRsharing.org. https://www.gwascentral.org/.

Beck, Tim, Tom Shorter, and Anthony J. Brookes. "GWAS Central: A Comprehensive Resource for the Discovery and Comparison of Genotype and Phenotype Data from Genome-Wide Association Studies." *Nucleic Acids Research* 48, no. D1 (08 2020): D933–40. https://doi.org/10.1093/nar/gkz895.

Bedagkar-Gala, Apurva, and Shishir K Shah. "A Survey of Approaches and Trends in Person Re-Identification." *Image and Vision Computing* 32, no. 4 (April 2014): 270–86. https://doi.org/10.1016/j.imavis.2014.02.001.

Bedford, Trevor, Richard Neher, James Hadfield, Emma Hodcroft, Thomas Sibley, John Huddleston, Jover Lee, et al. "Nextstrain Genomic epidemiology of novel coronavirus," 2020. https://nextstrain.org/ncov/global.

Benson, Dennis A., Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. "GenBank." *Nucleic Acids Research* 41, no. Database issue (January 2013): D36-42. https://doi.org/10.1093/nar/gks1195.

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.

Bernier, Alexander, and Adrian Thorogood. "Sharing Bioinformatic Data for Machine Learning: Maximizing Interoperability through License Selection." In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOINFORMATICS*, 3:226–32. Valletta, Malta, 2020. https://dx.doi.org/10.5220/0009179502260232.

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 CoE. "BioExcel Center of Excellence in Support of COVID-19 Research." BioExcel - Centre of Excellence for Computation Biomolecular Research, 2015. https://bioexcel.eu/bioexcel-center-of-excellence-in-support-of-covid-19-research/.

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/.

Blacketer, Margaret S, Erica A Voss, and Patrick B Ryan. "Applying the OMOP Common Data Model to Survey Data." *Medical Care*, 2013, S45–52.

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.

Bochove, Kees van, Emma Vos, Anne van Winzum, Julia Kurps, and Maxim Moinat. "Implementing FAIR in OHDSI: Challenges and Opportunities for EHDEN," 2020, 1.

Bonini, Sergio, Dawei Lin, Andrea Jackson Dipina, and Anne Cambon-Thomsen. "RDA-COVID19-Clinical." RDA, March 30, 2020. https://www.rd-alliance.org/groups/rda-covid19-clinical.

Boyd, Danah, and Kate Crawford. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5 (2012): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

BPRC. "IProX - Integrated Proteome Resources." IProX - integrated Proteome resources. Beijing Proteome Research Center (BPRC), 2019. https://www.iprox.org/.

Bradley, Declan Terence, Mariam Abdulmonem Mansouri, Frank Kee, and Leandro Martin Totaro Garcia. "A Systems Approach to Preventing and Responding to COVID-19." *EClinicalMedicine* 21 (April 1, 2020). https://doi.org/10.1016/j.eclinm.2020.100325.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, et al. "Minimum Information about a Microarray Experiment (MIAME)-toward Standards for Microarray Data." *Nature Genetics* 29, no. 4 (December 2001): 365–71. https://doi.org/10.1038/ng1201-365.

Brickley, Dan, et al., and et al. "Organization of Schemas," 2020. https://schema.org/docs/schemas.html.

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.

Bullock, H.E., L.L. Harlow, and S.A. Mulaik. "Causation Issues in Structural Equation Modeling Research." *Structural Equation Modeling: A Multidisciplinary Journal* 1, no. 3 (November 3, 2009): 253–67. https://doi.org/10.1080/10705519409539977.

Burke, R.L., K.C. Kronmann, C.C. Daniels, M. Meyers, D.K. Byarugaba, E. Dueger, T.A. Klein, B.P. Evans, and K.G. Vest. "A Review of Zoonotic Disease Surveillance Supported by the Armed Forces Health Surveillance Center." *Zoonoses and Public Health* 59, no. 3 (2012): 164–75. https://doi.org/10.1111/j.1863-2378.2011.01440.x.

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.

Calibr at Scripps Research. "ReframeDB: Open & Extendable Drug Repurposing Data." reframeDB, 2018. https://reframedb.org/.

Calisher, Charles H., James E. Childs, Hume E. Field, Kathryn V. Holmes, and Tony Schountz. "Bats: Important Reservoir Hosts of Emerging Viruses." *Clinical Microbiology Reviews* 19, no. 3 (July 1, 2006): 531–45. https://doi.org/10.1128/CMR.00017-06.

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.

Carr, David. "Coronavirus (COVID-19): Sharing Research Data | Wellcome," January 31, 2020. https://wellcome.ac.uk/coronavirus-covid-19/open-data.

———. "Publishers Make Coronavirus (COVID-19) Content Freely Available and Reusable | Wellcome," March 16, 2020. https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable.

Case, Nicky. "Protecting Lives & Liberty: How Contact Tracing Can Foil COVID-19 & Big Brother," 2020. https://ncase.me/contact-tracing/.

Caswell, Thomas, Sylvain Corlay, Romain François, Ralf Gommers, Alexandre Gramfort, Olivier Grisel, Jason Grout, et al. "COVID-19 Open Source Help Desk," April 30, 2020. https://covid-oss-help.org/.

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.

———. "National Pandemic Strategy," December 4, 2018. https://www.cdc.gov/flu/pandemic-resources/national-strategy/index.html.

———. "Person Under Investigation (PUI) and Case Report F.Pdf," April 23, 2020, 2.

———. "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.

———. "Weekly Provisional Death Counts from Death Certificate Data: COVID19, Pneumonia, Flu." Dataset. Center for Disease Control, USA, 2020. https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm.

———. "Zoonotic Diseases - One Health," February 19, 2020. https://www.cdc.gov/onehealth/basics/zoonotic-diseases.html.

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

Center for Computaitonal Mass Spectrometry. "MassIVE." Welcome to MassIVE, April 7, 2020. https://massive.ucsd.edu/.

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

CERN, and OpenAire. "Zenodo Digital Archive," 2020. https://zenodo.org/.

CESSDA. "CESSDA Vocabularies." Consortium of European Social Science Data Archives, 2019. https://vocabularies.cessda.eu/#!discover.

CESSDA Training Team. "Data Management Expert Guide." Data Management Expert Guide - CESSDA TRAINING, 2019. https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide.

CGH. "Global Health Security Agenda: GHSA Zoonotic Disease Action Package (GHSA Action Package Prevent-2)." Center for Global Health, January 16, 2019. https://www.cdc.gov/globalhealth/security/actionpackages/zoonotic_disease.htm.

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.

Chan, An-Wen, Jennifer M. Tetzlaff, Douglas G. Altman, Andreas Laupacis, Peter C. Gøtzsche, Karmela Krleža-Jerić, Asbjørn Hróbjartsson, et al. "SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials." *Annals of Internal Medicine* 158, no. 3 (February 5, 2013): 200–207. https://doi.org/10.7326/0003-4819-158-3-201302050-00583.

Chan, Jasper Fuk-Woo, Kelvin Kai-Wang To, Herman Tse, Dong-Yan Jin, and Kwok-Yung Yuen. "Interspecies Transmission and Emergence of Novel Viruses: Lessons from Bats and Birds." *Trends in Microbiology* 21, no. 10 (October 1, 2013): 544–55. https://doi.org/10.1016/j.tim.2013.05.005.

Chan Zuckerberg Initiative. "CZI Launches Funding Opportunity for Open Source Software." *Chan Zuckerberg Initiative* (blog), April 30, 2020. https://chanzuckerberg.com/rfa/essential-open-source-software-for-science/.

Chang, Christopher. "File Format Reference - PLINK 1.9," 2020. https://www.cog-genomics.org/plink2/formats#ped.

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.

Chen, N., M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, et al. "Epidemiological and Clinical Characteristics of 99 Cases of 2019 Novel Coronavirus Pneumonia in Wuhan, China: A Descriptive Study." *The Lancet* 395, no. 10223 (2020): 507–13. https://doi.org/10.1016/S0140-6736(20)30211-7.

Cheng, Vincent C. C., Shuk-Ching Wong, Jonathan H. K. Chen, Cyril C. Y. Yip, Vivien W. M. Chuang, Owen T. Y. Tsang, Siddharth Sridhar, Jasper F. W. Chan, Pak-Leung Ho, and Kwok-Yung Yuen. "Escalating Infection Control Response to the Rapidly Evolving Epidemiology of the Coronavirus Disease 2019 (COVID-19) Due to SARS-CoV-2 in Hong Kong." *INFECTION CONTROL AND HOSPITAL EPIDEMIOLOGY* 41, no. 5 (May 2020): 493–98. https://doi.org/10.1017/ice.2020.58.

Chomel, B.B., A. Belotto, and F.-X. Meslin. "Wildlife, Exotic Pets, and Emerging Zoonoses." *Emerging Infectious Diseases* 13, no. 1 (2007): 6–11. https://doi.org/10.3201/eid1301.060480.

Chowdhury, Rajiv, Kevin Heng, Md Shajedur Rahman Shawon, Gabriel Goh, Daisy Okonofua, Carolina Ochoa-Rosales, Valentina Gonzalez-Jaramillo, et al. "Dynamic Interventions to Control COVID-19 Pandemic: A Multivariate Prediction Modelling Study Comparing 16 Worldwide Countries." *European Journal of Epidemiology*, May 19, 2020. https://doi.org/10.1007/s10654-020-00649-w.

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.

Chue Hong, Neil, Martin Fenner, and Daniel S. Katz. "Software Citation Implementation Working Group." FORCE11, April 7, 2017. https://www.force11.org/group/software-citation-implementation-working-group.

Clark, Karen, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. "GenBank." *Nucleic Acids Research* 44, no. D1 (January 4, 2016): D67-72. https://doi.org/10.1093/nar/gkv1276.

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

Clobridge, Abby. "Building a Digital Repository Program with Limited Resources." In *Building a Digital Repository Program with Limited Resources*, edited by Abby Clobridge, 85–109. Chandos Information

Professional Series. Chandos Publishing, 2010. https://doi.org/10.1016/B978-1-84334-596-1.50005-5.

Clunie, David A. "DICOM Structured Reporting and Cancer Clinical Trials Results." *Cancer Informatics* 4 (2007): 33–56.

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.

COAR. "Recommendations for COVID-19 Resources in Repositories." Confederation of Open Access Repositories, 2020. https://www.coar-repositories.org/news-updates/covid19-recommendations/.

Coccia, Mario. "Two Mechanisms for Accelerated Diffusion of COVID-19 Outbreaks in Regions with High Intensity of Population and Polluting Industrialization: The Air Pollution-to-Human and Human-to-Human Transmission Dynamics." Cold Spring Harbor Laboratory Press, April 11, 2020. https://www.medrxiv.org/content/10.1101/2020.04.06.20055657v1.

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.

CODATA, Committee on Data of the International Science Council, CODATA International Data Policy Committee, CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice, Simon Hodson, Barend Mons, Paul Uhlir, and Lili Zhang. "The Beijing Declaration on Research Data," November 25, 2019. https://doi.org/10.5281/zenodo.3552330.

CODATA, RDC. "International Research Data Management Glossary (IRiDiuM)," 2017. https://codata.org/initiatives/working-groups/standard-glossary-for-research-data-management-iridium/.

Coffey, Barbara. "LibGuides: Finance: Info on Company Ids and Linking Data Sources," April 13, 2020. https://libguides.princeton.edu/c.php?g=939414&p=6776005.

Consultative Committee for Space Data Systems. "Audit and Certification of Trustworthy Digital Repositories." Consultative Committee for Space Data Systems, 2011. https://public.ccsds.org/pubs/652x0m1.pdf.

CoreTrustSeal. "Core Certified Repositories." *CoreTrustSeal* (blog), June 28, 2017. https://www.coretrustseal.org/why-certification/certified-repositories/.

———. "CoreTrustSeal." CoreTrustSeal, 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.

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.

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.

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

COVID19India. "Coronavirus in India: Latest Map and Case Count." Dataset, 2020. https://www.covid19india.org.

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

———. "Creative Commons — Attribution 4.0 International — CC BY 4.0," 2016. https://creativecommons.org/licenses/by/4.0/.

———. "Creative Commons — CC0 1.0 Universal," 2016. https://creativecommons.org/publicdomain/zero/1.0/.

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

Data Citation Synthesis Group. "Joint Declaration of Data Citation Principles - FINAL | FORCE11," 2014. https://doi.org/10.25490/a97f-egyk.

Data Documentation Initiative. "Controlled Vocabularies - Overview Table of Latest Versions | Data Documentation Initiative," 2018. https://ddialliance.org/controlled-vocabularies.

DataCite. "Re3data.Org: Registry of Research Data Repositories," 2020. https://doi.org/10.17616/R3D.

———. "Welcome to DataCite," 2020. https://datacite.org/.

Datacovid. "Barometer Covid19," 2020. https://datacovid.org/.

Datacovid.org. "COVID-19 Barometer," 2020. https://datacovid.org/.

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

Day, Michael J., Edward Breitschwerdt, Sarah Cleaveland, Umesh Karkare, Chand Khanna, Jolle Kirpensteijn, Thijs Kuiken, et al. "Surveillance of Zoonotic Infectious Disease Transmitted by Small Companion Animals - Volume 18, Number 12—December 2012 - Emerging Infectious Diseases Journal - CDC," 2012. https://doi.org/10.3201/eid1812.120664.

DCC. "DMP Online." Digital Curation Centre, 2010. https://dmponline.dcc.ac.uk/.

———. "Example DMPs and Guidance." Digital Curation Centre, 2020. http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples.

DDBJ. "DDBJ Annotated/Assembled Sequences," September 30, 2019. https://fairsharing.org/FAIRsharing.k337f0. https://www.ddbj.nig.ac.jp/.

———. "Japanese Genotype-Phenotype Archive (JGA)," February 21, 2018. https://fairsharing.org/FAIRsharing.pwgf4p. https://www.ddbj.nig.ac.jp/jga/index-e.html.

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.

DDBJ, Jun Mashima, Takehide Kosuge, and Osamu Ogasawara. "Genomic Expression Archive," July 25, 2018. https://fairsharing.org/. https://www.ddbj.nig.ac.jp/gea/index-e.html.

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

De Silva, De Silva, Bruna Galobardes, John Plummer, Eveline Herbst, Jim Todd, Chifundo Kanjala, Le Doare Le Doare, et al. "LMIC Covid Core Questionnaire," May 2020. https://wellcomecloud-my.sharepoint.com/:w:/g/personal/b_galobardes_wellcome_ac_uk/EX963zhZHWlPkYsDM79jQcwBoK4RD7JrYx8k4YFo-Ep6mA?rtime=k2_HF54b2Eg.

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.

Degeling, C., J. Johnson, M. Ward, A. Wilson, and G. Gilbert. "A Delphi Survey and Analysis of Expert Perspectives on One Health in Australia." *EcoHealth* 14, no. 4 (2017): 783–92. https://doi.org/10.1007/s10393-017-1264-7.

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.

[Di Cosmo], Roberto, Morane Gruenpeter, and Stefano Zacchiroli. "204.4 Identifiers for Digital Objects: The Case of Software Source Code Preservation.," September 21, 2018. https://doi.org/10.17605/OSF.IO/KDE56.

DICOM Standards Committee. "DICOM Standard," 2020. https://www.dicomstandard.org/.

———. "DICOMweb[TM] and Other Standards – DICOM Standard," 2020. https://www.dicomstandard.org/dicomweb/dicomweb-and-hl7-fhir/.

———. "Digital Imaging and Communications in Medicine (DICOM) Supplements," January 25, 2011. https://www.dicomstandard.org/supplements/.

———. "Supplement 142: Clinical Trial De-Identification Profiles," 2011. ftp://medical.nema.org/medical/dicom/final/sup142_ft.pdf.

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.

DORA. "San Francisco Declaration on Research Assessment," 2016. https://sfdora.org/read/.

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.

Dryad. "Dryad Home - Publish and Preserve Your Data," 2019. https://datadryad.org/stash.

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/.

Dublin Core. "DCMI: Dublin Core[TM]," May 21, 2020. https://dublincore.org/specifications/dublin-core/.

Duncan, George T., Mark Elliot, and Gonzalez Juan Jose Salazar. *Statistical Confidentiality: Principles and Practice*. Statistics for Social and Behavioral Sciences. New York: Springer-Verlag, 2011. https://doi.org/10.1007/978-1-4419-7802-8.

Duncan, M.A., D. Drociuk, A. Belflower-Thomas, D. Van Sickle, J.J. Gibson, C. Youngblood, and W.R. Daley. "Follow-Up Assessment of Health Consequences after a Chlorine Release from a Train Derailment-

Graniteville, SC, 2005." *Journal of Medical Toxicology* 7, no. 1 (2011): 85–91. https://doi.org/10.1007/s13181-010-0130-6.

Dunn, Patrick. "HIPC ImmPort Cytometry Data Standards." Online resource. figshare. National Institutes of Health, May 19, 2020. https://doi.org/10.35092/yhjc.12314180.v1.

———. "ImmPort Data Upload Workflow and Templates." Online resource. figshare. National Institutes of Health, May 19, 2020. https://doi.org/10.35092/yhjc.12311945.v1.

E13 Committee. "Specification for Analytical Data Interchange Protocol for Chromatographic Data." ASTM International, 2014. https://doi.org/10.1520/E1947-98R14.

eCDC. "Country Preparedness Plans on Zoonotic Influenza." European Centre for Disease Prevention and Control, 2015. https://www.ecdc.europa.eu/en/avian-influenza-humans/country-preparedness-plans-avian-influenza-humans.

ECDC. "COVID-19 Situation Update Worldwide, as of 9 June 2020." European Centre for Disease Prevention and Control, 2020. https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases.

———. "Geographic Distribution of COVID-19 Cases Worldwide," April 15, 2020. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide.

———. "How ECDC Collects and Processes COVID-19 Data." European Centre for Disease Prevention and Control, 2020. https://www.ecdc.europa.eu/en/covid-19/data-collection.

———. "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.

eCDC. "The European Union One Health 2018 Zoonoses Report." European Centre for Disease Prevention and Control, December 12, 2019. https://www.ecdc.europa.eu/en/publications-data/european-union-one-health-2018-zoonoses-report.

Edmunds, Rorie, Mary Vardigan, and Lesley Rickards. "Repository Audit and Certification DSA–WDS Partnership WG." RDA, May 21, 2014. https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html.

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.

———. "ELIXIR," 2020. https://elixir-europe.org/about-us.

———. "Galaxy-ELIXIR Webinar Series: FAIR Data and Open Infrastructures to Tackle the COVID-19 Pandemic | ELIXIR," April 30, 2020. https://elixir-europe.org/events/webinar-galaxy-elixir-covid19.

Elsevier. "SoftwareX," 2020. https://www.journals.elsevier.com/softwarex.

EMBL-EBI. "Annotare: Accepted Raw Microarray Files Formats." Accepted Raw Microarray Files Formats < Guide < Annotare < EMBL-EBI, 2016. https://www.ebi.ac.uk/fg/annotare/help/accepted_raw_ma_file_formats.html.

———. "ArrayExpress," 2020. https://fairsharing.org/FAIRsharing.6k0kwd. https://www.ebi.ac.uk/arrayexpress/.

———. "BioSamples," 2020. https://doi.org/10.25504/FAIRsharing.ewjdq6. https://www.ebi.ac.uk/biosamples/.

———. "European Nucleotide Archive (ENA)," 2020. https://fairsharing.org/FAIRsharing.dj8nt8. 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.

———. "MetaboLights - Metabolomics Experiments and Derived Information." MetaboLights - Metabolomics experiments and derived information, 2017. https://www.ebi.ac.uk/metabolights/.

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

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

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

EQUATOR. "Enhancing the QUAlity and Transparency Of health Research." The EQUATOR Network, 2020. https://www.equator-network.org/.

eScience Center. "FAIR Research Software." FAIR Research Software, 2020. https://fair-software.nl/recommendations/repository.

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 Clinical Research Infrastructure Network. "Clinical Research Metadata Repository | ECRIN," 2020. https://www.ecrin.org/clinical-research-metadata-repository.

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/node/2_en.

———. 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 Genome-phenome Archive (EGA). "The EGA European Genome-Phenome Archive," 2020. https://fairsharing.org/FAIRsharing.mya1ff. https://ega-archive.org/.

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.

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

eurostat. "NUTS - Nomenclature of Territorial Units for Statistics," 2019. https://ec.europa.eu/eurostat/web/nuts/background.

Everard, M., P. Johnston, D. Santillo, and C. Staddon. "The Role of Ecosystems in Mitigation and Management of Covid-19 and Other Zoonoses." *Environmental Science and Policy* 111 (2020): 7–17. https://doi.org/10.1016/j.envsci.2020.05.017.

Exaptive. "Cognitive City: COVID-19 Resources." Exaptive, and the Bill and Melinda Gates Foundation, 2020. https://covid-19.cognitive.city/cognitive/community/resource-gallery.

FAIR principles. "FAIR Research Software," 2020. https://fair-software.nl.

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

Falzon, L.C., L. Alumasa, F. Amanya, E. Kang'ethe, S. Kariuki, K. Momanyi, P. Muinde, et al. "One Health in Action: Operational Aspects of an Integrated Surveillance System for Zoonoses in Western Kenya." *Frontiers in Veterinary Science* 6 (2019). https://doi.org/10.3389/fvets.2019.00252.

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.

Fauci, Anthony S., and David M. Morens. "The Perpetual Challenge of Infectious Diseases." *New England Journal of Medicine* 366, no. 5 (February 2, 2012): 454–61. https://doi.org/10.1056/NEJMra1108296.

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.

Ferguson, N.M., D.A.T. Cummings, C. Fraser, J.C. Cajka, P.C. Cooley, and D.S. Burke. "Strategies for Mitigating an Influenza Pandemic." *Nature* 442, no. 7101 (2006): 448–52. https://doi.org/10.1038/nature04795.

FGED. "Minimal Information about a High Throughput SEQuencing Experiment." Functional Genomics Data Society, March 2008. https://fairsharing.org/FAIRsharing.a55z32. http://fged.org/projects/minseqe/.

FigShare. "FigShare Repository," 2020. https://figshare.com/.

Fineberg, H.V. "Global Health: Pandemic Preparedness and Response - Lessons from the H1N1 Influenza of 2009." *New England Journal of Medicine* 370, no. 14 (2014): 1335–42. https://doi.org/10.1056/NEJMra1208802.

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.

First Nations Information Governance Centre (FNIGC). "Home | FNIGC," 2020. https://fnigc.ca/.

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.

Floridi, Luciano. "Open Data, Data Protection, and Group Privacy." *Philosophy & Technology* 27, no. 1 (2014): 1–3. https://doi.org/10.1007/s13347-014-0157-8.

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

Foley, Desmond, Pollie Rueda, Richard Wilkerson, and the ESWG (Entomological Surveillance Working Group). "VectorMap: Know the Vector, Know the Threat." Walter Reed Biosystematics Unit (WRBU), U.S. Department of Defense, 2019. http://vectormap.si.edu/Project_ESWG.htm.

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

Franklin, A.B., and S.N. Bevins. "Spillover of SARS-CoV-2 into Novel Wild Hosts in North America: A Conceptual Model for Perpetuation of the Pathogen." *Science of the Total Environment* 733 (2020). https://doi.org/10.1016/j.scitotenv.2020.139358.

French COVID-19. "COVID-19 Funding Opportunities." *Reacting* (blog), March 12, 2020. https://reacting.inserm.fr/covid-19-funding-opportunities/.

Freymann, John B., Justin S. Kirby, John H. Perry, David A. Clunie, and C. Carl Jaffe. "Image Data Sharing for Biomedical Research--Meeting HIPAA Requirements for De-Identification." *Journal of Digital Imaging* 25, no. 1 (February 2012): 14–24. https://doi.org/10.1007/s10278-011-9422-x.

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.

Frontera, Antonio, Claire Martin, Kostantinos Vlachos, and Giovanni Sgubin. "Regional Air Pollution Persistence Links to COVID-19 Infection Zoning." *The Journal of Infection*, April 10, 2020. https://doi.org/10.1016/j.jinf.2020.03.045.

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/.

Gates, B. "Responding to Covid-19 - A Once-in-a-Century Pandemic?" *New England Journal of Medicine* 382, no. 18 (2020): 1677–79. https://doi.org/10.1056/NEJMp2003762.

Gebreyes, W.A., J. Dupouy-Camet, M.J. Newport, C.J.B. Oliveira, L.S. Schlesinger, Y.M. Saif, S. Kariuki, et al. "The Global One Health Paradigm: Challenges and Opportunities for Tackling Infectious Diseases at the Human, Animal, and Environment Interface in Low-Resource Settings." *PLoS Neglected Tropical Diseases* 8, no. 11 (2014). https://doi.org/10.1371/journal.pntd.0003257.

GECCO. "Covid-19 Research-Dataset - Datasets," May 20, 2020. https://art-decor.org/art-decor/decor-datasets--covid19f-?id=&effectiveDate=&conceptId=&conceptEffectiveDate=.

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://www.ratswd.de/en/publication/output-series/2855.

German National Cohort. "NAKO Gesundheitsstudie - Kontakt," March 31, 2020. https://nako.de/allgemeines/kontakt/.

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

Gershman, Boris, David P. Guo, and Issa J. Dahabreh. "Using Observational Data for Personalized Medicine When Clinical Trial Evidence Is Limited." *Fertility and Sterility* 109, no. 6 (2018): 946–51. https://doi.org/10.1016/j.fertnstert.2018.04.005.

GESIS Panel Team. "GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in GermanyGESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany." GESIS Data Archive, 2020. https://doi.org/10.4232/1.13520.

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.

GHRU. "GHRU - COVID Questionnaire - v6.Docx." Dropbox, March 1, 2020. https://www.dropbox.com/s/auvvey4utibd85s/GHRU%20-%20COVID%20Questionnaire%20-%20v6.docx?dl=0.

Giacomoni, F., G. Le Corguille, M. Monsoor, M. Landi, P. Pericard, M. Petera, C. Duperier, et al. "Workflow4Metabolomics: A Collaborative Research Infrastructure for Computational Metabolomics." *Bioinformatics* 31, no. 9 (May 1, 2015): 1493–95. https://doi.org/10.1093/bioinformatics/btu813.

GIDA. "Global Indigenous Data Alliance." Global Indigenous Data Alliance, 2020. https://www.gida-global.org.

GigaScience. "Instructions to Authors." Oxford Academic, 2020. https://academic.oup.com/gigascience/pages/instructions_to_authors.

GigitalReach. "Should We Be Worried about a Tracing App during the COVID-19?" *DigitalReach* (blog), May 15, 2020. https://digitalreach.asia/should-we-be-worried-about-a-tracing-app-during-the-covid-19/.

Giordano, Giulia, Franco Blanchini, Raffaele Bruno, Patrizio Colaneri, Alessandro Di Filippo, Angela Di Matteo, and Marta Colaneri. "Modelling the COVID-19 Epidemic and Implementation of Population-Wide Interventions in Italy." *Nature Medicine*, April 22, 2020, 1–6. https://doi.org/10.1038/s41591-020-0883-7.

GitHub. "Build Software Better, Together." GitHub, 2020. https://github.com.

———. "Choose an Open Source License," 2020. https://choosealicense.com/.

———. "Making Your Code Citable · GitHub Guides," October 2016. https://guides.github.com/activities/citable-code/.

———. "SAMtools: Hts-Specs." 2012. Reprint, samtools, May 13, 2020. https://fairsharing.org/FAIRsharing.cfzz0h. https://github.com/samtools/hts-specs.

GitHub Inc. "Choose an Open Source License." Choose a License, 2020. https://choosealicense.com/.

———. "GitHub." GitHub, 2020. https://github.com/.

GitLab. "The First Single Application for the Entire DevOps Lifecycle - GitLab | GitLab," 2020. https://about.gitlab.com/.

GLEWS. "The Global Early Warning System for Health Threats and Emerging Risks at the Human–Animal–Ecosystems Interface," 2006. http://www.glews.net/.

Global Alliance for Genomics & Health. "Framework for Responsible Sharing of Genomic and Health-Related Data," December 9, 2014. 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 and Health (GA4GH. "CRAM," 2020. https://www.ga4gh.org/cram/.

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/.

Global outbreak alert and response network. "COVID-19 Knowledge Hub | GOARN," 2020. https://extranet.who.int/goarn/COVID19Hub.

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.

GO FAIR. "FAIR Principles." GO FAIR, 2017. https://www.go-fair.org/fair-principles/.

Goldberg, Mark, and Paul Villeneuve. "Air Pollution, COVID-19 and Death: The Perils of Bypassing Peer Review." The Conversation, 2020. http://theconversation.com/air-pollution-covid-19-and-death-the-perils-of-bypassing-peer-review-136376.

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/.

———. "Google LLC 'COVID-19 Community Mobility Report.'" COVID-19 Community Mobility Report, 2020. https://www.google.com/covid19/mobility?hl=en.

Gorbalenya, A.E., S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B.L. Haagmans, et al. "The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-NCoV and Naming It SARS-CoV-2." *Nature Microbiology* 5, no. 4 (2020): 536–44. https://doi.org/10.1038/s41564-020-0695-z.

Gostin, L.O., and R. Katz. "The International Health Regulations: The Governing Framework for Global Health Security." *Milbank Quarterly* 94, no. 2 (2016): 264–313. https://doi.org/10.1111/1468-0009.12186.

Government of the Democratic Republic of the Congo, and World Health Organization. "Strategic_response_plan.Pdf," 2019. https://www.un.org/ebolaresponsedrc/sites/www.un.org.ebolaresponsedrc/files/strategic_response_plan.pdf.

Grange, Elisha S., Eric J. Neil, Michelle Stoffel, Angad P. Singh, Ethan Tseng, Kelly Resco-Summers, B. Jane Fellner, et al. "Responding to COVID-19: The UW Medicine Information Technology Services Experience." *APPLIED CLINICAL INFORMATICS* 11, no. 2 (March 2020): 265–75. https://doi.org/10.1055/s-0040-1709715.

Greenfield, Jay, Rajini Nagrani, Meg Sears, Claire C Austin, and the RDA-COVID19-WG. "A Full Spectrum View of the COVID-19 Data Domain: An Epidemiological Data Model." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

Greenfield, Jay, Meg Sears, Rajini Nagrani, Gary Mazzaferro, Anna Widyastuti, Claire C Austin, and the RDA-COVID19-WG. "Common Data Models and Full Spectrum Epidemiology: Epi-STACK Architecture for COVID-19 Epidemiology Datasets." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

Greenfield, Jay, E. Z. Tonnang, Gary Mazzaferro, Claire C Austin, and the RDA-COVID19-WG. "Epi-TRACS: Rapid Detection and Whole System Response for Emerging Pathogens Such as SARS-CoV-2 Virus and the COVID-19 Disease That It Causes." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

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," 2003. https://fairsharing.org/FAIRsharing.sggb1n. https://mblab.wustl.edu/GTF22.html.

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

GWAS Catalog. "GWAS Catalog," May 17, 2020. https://www.ebi.ac.uk/gwas/.

Haesler, Barbara, William Gilbert, Bryony Anne Jones, Dirk Udo Pfeiffer, Jonathan Rushton, and Martin Joachim Otte. "The Economic Value of One Health in Relation to the Mitigation of Zoonotic Disease Risks." Edited by Mackenzie, JS and Jeggo, M and Daszak, P and Richt, JA. *ONE HEALTH: THE HUMAN-ANIMAL-ENVIRONMENT INTERFACES IN EMERGING INFECTIOUS DISEASES: THE CONCEPT AND EXAMPLES OF A ONE HEALTH APPROACH*. Current Topics in Microbiology and Immunology. HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY: SPRINGER-VERLAG BERLIN, 2013. https://doi.org/10.1007/82_2012_239.

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.

Harvard. "COVID-19 Hospital Capacity Estimates 2020." Harvard Global Health Institute, 2020. https://globalepidemics.org/.

Harvard Dataverse. "Harvard Dataverse," 2020. https://dataverse.harvard.edu/.

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, Daniel. "Protecting Groups from Genetic Research." *Bioethics* 22, no. 3 (March 2008): 157–165. https://doi.org/10.1111/j.1467-8519.2007.00625.x.

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/764
0426.

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

Healy, Kieran. *Rpackage (Covdata) - COVID19 Case and Mortality Time Series*, 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.

Hernán, Miguel A., and James M. Robins. "Using Big Data to Emulate a Target Trial When a Randomized
Trial Is Not Available." *American Journal of Epidemiology* 183, no. 8 (15 2016): 758–64.
https://doi.org/10.1093/aje/kwv254.

HL7. "Fast Healthcare Interoperability Resources Standard." HL7®, November 1, 2019.
https://www.hl7.org/fhir/.

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

———. "Summary - FHIR v4.0.1," November 1, 2019. https://hl7.org/FHIR/summary.html.

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.

Holshue, Michelle L., Chas DeBolt, Scott Lindquist, Kathy H. Lofy, John Wiesman, Hollianne Bruce,
Christopher Spitters, et al. "First Case of 2019 Novel Coronavirus in the United States." *NEW
ENGLAND JOURNAL OF MEDICINE*. WALTHAM WOODS CENTER, 860 WINTER ST,, WALTHAM, MA
02451-1413 USA: MASSACHUSETTS MEDICAL SOC, March 5, 2020.
https://doi.org/10.1056/NEJMoa2001191.

Homeland Security Council. "Pandemic-Influenza-Implementation.Pdf," May 2006.
https://www.cdc.gov/flu/pandemic-resources/pdf/pandemic-influenza-implementation.pdf.

Horai, Hisayuki, Masanori Arita, Shigehiko Kanaya, Yoshito Nihei, Tasuku Ikeda, Kazuhiro Suwa, Yuya
Ojima, et al. "MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences."
Journal of Mass Spectrometry 45, no. 7 (2010): 703–14. https://doi.org/10.1002/jms.1777.

Howison, James, and Julia Bullard. "Software in the Scientific Literature: Problems with Seeing, Finding,
and Using Software Mentioned in the Biology Literature." *Journal of the Association for Information
Science and Technology* 67, no. 9 (2016): 2137–55. https://doi.org/10.1002/asi.23538.

Hu, Tao, Weihe Wendy Guan, Xinyan Zhu, Yuanzheng Shao, Lingbo Liu, Jing Du, Hongqiang Liu, et al.
"Building an Open Resources Repository for COVID-19 Research." SSRN Scholarly Paper. Rochester,
NY: Social Science Research Network, April 8, 2020. https://doi.org/10.2139/ssrn.3587704.

Hughes, James M., Mary E. Wilson, Brian L. Pike, Karen E. Saylors, Joseph N. Fair, Matthew LeBreton, Ubald
Tamoufe, Cyrille F. Djoko, Anne W. Rimoin, and Nathan D. Wolfe. "The Origin and Prevention of
Pandemics." *Clinical Infectious Diseases* 50, no. 12 (June 15, 2010): 1636–40.
https://doi.org/10.1086/652860.

HUPO Proteomics Standards Initiative. "MzIdentML | HUPO Proteomics Standards Initiative," March
2017. http://www.psidev.info/mzidentml.

———. "MzQuantML | HUPO Proteomics Standards Initiative," February 2013.
http://www.psidev.info/mzquantml.

HUPO PSI. "GelML 1.1.0 Specification." GelML 1.1.0 Specification | HUPO Proteomics Standards Initiative,
June 2010. http://www.psidev.info/gelml/1.0.

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

———. "MzTab Specification 1.0.0." mzTab Specification 1.0.0 | HUPO Proteomics Standards Initiative, June 2014. http://www.psidev.info/mztab.

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

———. "TraML 1.0.0 Specification." TraML 1.0.0 Specification | HUPO Proteomics Standards Initiative, April 22, 2013. http://www.psidev.info/traml.

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 Council on Archives, and International Conference of Information Commissioners. "COVID-19: The Duty to Document Does Not Cease in a Crisis, It Becomes More Essential - Digital Preservation Coalition," May 4, 2020. https://www.ica.org/sites/default/files/covid_the_duty_to_document_is_essential.pdf.

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 Severe Acute Respiratory and Emerging Infection Consortium, and World Health Organization. "ISARIC_COVID-19_RAPID_CRF_24MAR20_EN.Pdf," March 24, 2020. https://media.tghn.org/medialibrary/2020/04/ISARIC_COVID-19_RAPID_CRF_24MAR20_EN.pdf.

———. "ISARIC_WHO_nCoV_CORE_CRF_23APR20.Pdf," April 23, 2020. https://media.tghn.org/medialibrary/2020/05/ISARIC_WHO_nCoV_CORE_CRF_23APR20.pdf.

International Union of Crystallography. "Catalogue of Metadata Resources for Crystallographic and Related Applications." Crystallographic data, 2014. https://www.iucr.org/resources/data/dddwg/metadata-catalogue.

———. "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.

Inter-university Consortium for Political and Social Research. "COVID-19 Data Repository." Dataset. Inter-university Consortium for Political and Social Research, 2020. https://www.openicpsr.org/openicpsr/covid19.

ISA. "Standardizing Metadata for Scientific Experiments." ISA tools, March 2, 2018. https://isa-tools.org/.

ISAC. "Flow Cytometry Data File Standard." International Society for Advancement of Cytometry (ISAC), 2008. https://fairsharing.org/FAIRsharing.qrr33y. https://isac-net.org/page/Data-Standards?&hhsearchterms=%22flow+and+cytometry+and+data+and+file+and+standard%22.

———. "Minimum Information about Flow Cytometry." International Society for Advancement of Cytometry (ISAC), 2015. fairsharing.org. https://isac-net.org/page/MIFlowCyt.

ISARIC. "COVID-19 Clinical Research Resources." International Severe Acute Respiratory and emerging INfection Consortium - The Global Health Network, 2020. https://infograph.venngage.com/pe/Mxg90X1jTc?border=false.

ISO. "ISO - ISO 3166 — Country Codes." ISO, March 25, 2020. https://www.iso.org/iso-3166-country-codes.html.

ISO/TC 211. "ISO 19115-1:2014." ISO, 2014. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/37/53798.html.

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.

Ivanov, Dmitry. "Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-Based Analysis on the Coronavirus Outbreak (COVID-19/SARS-CoV-2) Case." *TRANSPORTATION RESEARCH PART E-LOGISTICS AND TRANSPORTATION REVIEW* 136 (April 2020). https://doi.org/10.1016/j.tre.2020.101922.

Jackson, James K, Martin A Weiss, Andres B Schwarzenberg, and Rebecca M Nelson. "Global Economic Effects of COVID-19." *Congressional Research Service*, 2020, 2020-05–15.

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* 115, no. 42 (October 16, 2018): 10750–55. https://doi.org/10.1073/pnas.1810137115.

Jee, Youngmee. "WHO International Health Regulations Emergency Committee for the COVID-19 Outbreak." *EPIDEMIOLOGY AND HEALTH* 42 (March 19, 2020). https://doi.org/10.4178/epih.e2020013.

Jeffery, Keith, Rebecca Koskela, and Alex Ball. "Metadata IG." RDA, May 24, 2013. https://www.rd-alliance.org/groups/metadata-ig.html.

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

Jiménez, Rafael C., Mateusz Kuzak, Monther Alhamdoosh, Michelle Barker, Bérénice Batut, Mikael Borg, Salvador Capella-Gutierrez, et al. "Four Simple Recommendations to Encourage Best Practices in Research Software." *F1000Research* 6 (2017): 876. https://doi.org/10.12688/f1000research.11407.1.

JOSS. "Journal of Open Source Software," 2020. https://joss.theoj.org.

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.

Kanjala, C. "Provenance of 'after the Fact' Harmonised Community-Based Demographic and HIV Surveillance Data from ALPHA Cohorts." Doctoral, London School of Hygiene & Tropical Medicine, 2020. https://doi.org/Kanjala, C <http://researchonline.lshtm.ac.uk/view/creators/ecpsckan.html>; (2020) Provenance of "after the fact" harmonised community-based demographic and HIV surveillance data from ALPHA cohorts. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: https://doi.org/10.17037/PUBS.04655994 <https://doi.org/10.17037/PUBS.04655994>.

Karesh, W.B., A. Dobson, J.O. Lloyd-Smith, J. Lubroth, M.A. Dixon, M. Bennett, S. Aldrich, et al. "Ecology of Zoonoses: Natural and Unnatural Histories." *The Lancet* 380, no. 9857 (2012): 1936–45. https://doi.org/10.1016/S0140-6736(12)61678-X.

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.

Kherbst, and Eehlers. "COVID-19 Screening Form 2020.pdf." Dropbox, April 27, 2020. https://www.dropbox.com/s/iuhe4366msdfq5c/COVID-19%20Screening%20Form%202020.pdf?dl=0.

Kilpatrick, A.M., and S.E. Randolph. "Drivers, Dynamics, and Control of Emerging Vector-Borne Zoonotic Diseases." *The Lancet* 380, no. 9857 (2012): 1946–55. https://doi.org/10.1016/S0140-6736(12)61151-9.

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.

Kissler, Stephen M., Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. "Projecting the Transmission Dynamics of SARS-CoV-2 through the Postpandemic Period." *Science* 368, no. 6493 (May 22, 2020): 860–68. https://doi.org/10.1126/science.abb5793.

Klyne, Graham, Jeremy J. Carroll, Pat Haye, Sergey Melnik, and Patrick Stickler. "Resource Description Framework (RDF): Concepts and Abstract Syntax," 2004. https://www.w3.org/TR/rdf-concepts/.

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.

Kukutai, Tahu, and John Taylor. "Data Sovereignty for Indigenous Peoples: Current Practice and Future Needs." In *Indigenous Data Sovereignty*, edited by Tahu Kukutai and John Taylor, 1st ed. ANU Press, 2016. https://doi.org/10.22459/CAEPR38.11.2016.01.

Kurt, Ozlem Kar, Jingjing Zhang, and Kent E Pinkerton. "Pulmonary Health Effects of Air Pollution." *Current Opinion in Pulmonary Medicine* 22, no. 2 (March 2016): 138–43. https://doi.org/10.1097/MCP.0000000000000248.

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.

Kushida, Clete A., Deborah A. Nichols, Rik Jadrnicek, Ric Miller, James K. Walsh, and Kara Griffin. "Strategies for De-Identification and Anonymization of Electronic Health Record Data for Use in Multicenter Research Studies." *Medical Care* 50, no. Suppl (July 2012): S82-101. https://doi.org/10.1097/MLR.0b013e3182585355.

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

Laine, Jessica E., and Oliver Robinson. "Framing Fetal and Early Life Exposome Within Epidemiology." In *Unraveling the Exposome: A Practical View*, edited by Sonia Dagnino and Anthony Macherone, 87–123. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-319-89321-1_4.

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

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, Benjamin D. "Ten Simple Rules for Documenting Scientific Software." *PLOS Computational Biology* 14, no. 12 (December 20, 2018): e1006561. https://doi.org/10.1371/journal.pcbi.1006561.

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.

Leebens-Mack, Jim, Todd Vision, Eric Brenner, John E. Bowers, Steven Cannon, Mark J. Clement, Clifford W. Cunningham, et al. "Taking the First Steps towards a Standard for Reporting on Phylogenies: Minimum Information about a Phylogenetic Analysis (MIAPA)." *OMICS: A Journal of Integrative Biology* 10, no. 2 (June 1, 2006): 231–37. https://doi.org/10.1089/omi.2006.10.231.

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. "SAMtools: The Sequence Alignment/Map (SAM)." *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, 2019. https://www.loc.gov/preservation/resources/rfs/TOC.html.

Lin, Dawei, Jonathan Crabtree, Ingrid Dillo, Robert R. Downs, Rorie Edmunds, David Giaretta, Marisa De Giusti, et al. "The TRUST Principles for Digital Repositories." *Scientific Data* 7, no. 1 (May 14, 2020): 144. https://doi.org/10.1038/s41597-020-0486-7.

Lipidomics Standards Initiative. "Guidelines - Lipidomics-Standards-Initiative (LSI)," 2020. https://lipidomics-standards-initiative.org/guidelines.

LONIC. "LOINC." LOINC, December 13, 2019. https://loinc.org/.

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

Luis, Angela D., David T. S. Hayman, Thomas J. O'Shea, Paul M. Cryan, Amy T. Gilbert, Juliet R. C. Pulliam, James N. Mills, et al. "A Comparison of Bats and Rodents as Reservoirs of Zoonotic Viruses: Are Bats Special?" *Proceedings of the Royal Society B: Biological Sciences* 280, no. 1756 (April 7, 2013): 20122753. https://doi.org/10.1098/rspb.2012.2753.

Luke, D.A., and K.A. Stamatakis. "Systems Science Methods in Public Health: Dynamics, Networks, and Agents," Annual Review of Public Health, 33 (2012): 357–76. https://doi.org/10.1146/annurev-publhealth-031210-101222.

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.

MacFarlane, D., and R. Rocha. "Guidelines for Communicating about Bats to Prevent Persecution in the Time of COVID-19." *Biological Conservation* 248 (2020). https://doi.org/10.1016/j.biocon.2020.108650.

Mackintosh, Kate. "The Principles of Humanitarian Action in International Humanitarian Law. HPG Report." ODI, 2000. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/305.pdf.

Madden, Richard, Per Axelsson, Tahu Kukutai, Kalinda Griffiths, Christina Storm Mienna, Ngaire Brown, Clare Coleman, and Ian Ring. "Statistics on Indigenous Peoples: International Effort Needed." *Statistical Journal of the IAOS* 32, no. 1 (January 1, 2016): 37–41. https://doi.org/10.3233/SJI-160975.

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.

Mahmood, Sultan, Khaled Hasan, Michelle Colder Carras, and Alain Labrique. "Global Preparedness Against COVID-19: We Must Leverage the Power of Digital Health." *JMIR PUBLIC HEALTH AND SURVEILLANCE* 6, no. 2 (June 2020): 226–32. https://doi.org/10.2196/18980.

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.

Mantelero, Alessandro. "Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection." *Computer Law & Security Review* 32, no. 2 (April 2016): 238–55. https://doi.org/10.1016/j.clsr.2016.01.014.

Martelletti, Luigi, and Paolo Martelletti. "Air Pollution and the Novel Covid-19 Disease: A Putative Disease Risk Factor." *Sn Comprehensive Clinical Medicine*, April 15, 2020, 1–5. https://doi.org/10.1007/s42399-020-00274-4.

Martínez Cobo, José. "Martínez Cobo Study | United Nations For Indigenous Peoples," 1981. https://www.un.org/development/desa/indigenouspeoples/publications/martinez-cobo-study.html.

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.

Mavragani, Amaryllis. "Tracking COVID-19 in Europe: Infodemiology Approach." *JMIR PUBLIC HEALTH AND SURVEILLANCE* 6, no. 2 (June 2020): 233–45. https://doi.org/10.2196/18941.

McCandless, David. "COVID-19 CoronaVirus Infographic Datapack." Information is Beautiful, 2020. https://informationisbeautiful.net/visualizations/covid-19-coronavirus-infographic-datapack/.

McMichael, T.M., D.W. Currie, S. Clark, S. Pogosjans, M. Kay, N.G. Schwartz, J. Lewis, et al. "Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington." *The New England Journal of Medicine*, March 27, 2020. https://doi.org/10.1056/NEJMoa2005412.

Mental Health Europe. "Vulnerable Groups Should Be Protected during the COVID-19 Pandemic." Mental Health Europe, April 14, 2020. https://www.mhe-sme.org/vulnerable-groups-should-be-protected-during-the-covid-19-pandemic/.

Metabolomics Standards Initiative. "Metabolomics Standards Initiative (MSI) - Biological Sample Context WG," August 31, 2005. http://msi-workgroups.sourceforge.net/bio-metadata.

Metadata Standards Catalog. "Metadata Standards Catalog," 2020. https://rdamsc.bath.ac.uk/.

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/.

Molecular Sciences Software Institute. "MolSSI – The Molecular Sciences Software Institute." MolSSI – The Molecular Sciences Software Institute, 2016. https://molssi.org/.

Molyneux, D., Z. Hallaj, G.T. Keusch, D.P. McManus, H. Ngowi, S. Cleaveland, P. Ramos-Jimenez, et al. "Zoonoses and Marginalised Infectious Diseases of Poverty: Where Do We Stand?" *Parasites and Vectors* 4, no. 1 (2011). https://doi.org/10.1186/1756-3305-4-106.

Morgan, Daniel, and John F Sargent. "Effects of COVID-19 on the Federal Research and Development Enterprise." CRS Reports. Washington, D.C.: Congressional Research Service (CRS), United States Library of Congress, April 10, 2020. https://crsreports.congress.gov/product/pdf/R/R46309.

Morse, S.S., J.A.K. Mazet, M. Woolhouse, C.R. Parrish, D. Carroll, W.B. Karesh, C. Zambrana-Torrelio, W.I. Lipkin, and P. Daszak. "Prediction and Prevention of the next Pandemic Zoonosis." *The Lancet* 380, no. 9857 (2012): 1956–65. https://doi.org/10.1016/S0140-6736(12)61684-5.

Mozilla. "MOSS Launches COVID-19 Solutions Fund." The Mozilla Blog, April 30, 2020. https://blog.mozilla.org/blog/2020/03/31/moss-launches-covid-19-solutions-fund.

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.

MS-DIAL. "CompMS | MS-DIAL," January 2019. http://prime.psc.riken.jp/compms/msdial/main.html#MSP.

Munthali, George N. Chidimbah, and Wu Xuelian. "Covid-19 Outbreak on Malawi Perspective." *ELECTRONIC JOURNAL OF GENERAL MEDICINE* 17, no. 4 (2020). https://doi.org/10.29333/ejgm/7871.

NASEM. "Achieving Sustainable Global Capacity for Surveillance and Response to Emerging Diseases of Zoonotic Origin: Workshop Summary." National Academies of Sciences, Engineering, and Medicine, December 31, 2008. https://doi.org/10.17226/12522.

———. "Coronavirus Resources Collection." National Academies of Sciences, Engineering, and Medicine, 2020. http://www.nap.edu/collection/94/coronavirus-resources.

———. "Emerging Viral Diseases: The One Health Connection: Workshop Summary." National Academies of Sciences, Engineering, and Medicine, 2015. https://www.nap.edu/catalog/18975/emerging-viral-diseases-the-one-health-connection-workshop-summary.

———. "Evaluating Data Types: A Guide for Decision Makers Using Data to Understand the Extent and Spread of COVID-19." National Academies of Sciences, Engineering, and Medicine, June 11, 2020. https://doi.org/10.17226/25826.

———. "Exploring Lessons Learned from Partnerships to Improve Global Health and Safety: Workshop in Brief." National Academies of Sciences, Engineering, and Medicine, June 7, 2018. https://doi.org/10.17226/21690.

———. "Infectious Disease Movement in a Borderless World: Workshop Summary." National Academies of Sciences, Engineering, and Medicine, 2010. https://www.nap.edu/download/12758.

———. "Rapid Expert Consultation on Data Elements and Systems Design for Modeling and Decision Making for the COVID-19 Pandemic (March 21, 2020)." National Academies of Sciences, Engineering, and Medicine, March 22, 2020. https://doi.org/10.17226/25755.

———. "Rapid Expert Consultations on the COVID-19 Pandemic: March 14, 2020-April 8, 2020." National Academies of Sciences, Engineering, and Medicine, April 30, 2020. National Academies of Sciences, Engineering, and Medicine. https://doi.org/10.17226/25784.

———. *Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases*. National Academies of Sciences, Engineering, and Medicine, 2009. https://doi.org/10.17226/12625.

NASEM, Gerald T. Keusch, Marguerite Pappaioanou, Mila C. Gonzalez, Kimberly A. Scott, and Peggy Tsai. *Achieving an Effective Zoonotic Disease Surveillance System*. *Sustaining Global Surveillance and Response to Emerging Zoonotic Diseases*. National Academies of Sciences, Engineering, and Medicine, 2009. https://www.ncbi.nlm.nih.gov/books/NBK215315/.

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," 2013. https://www.niaid.nih.gov/research/data-sharing-and-release-guidelines.

National Metabolomics Data Repository(NMDR). "Metabolomics Workbench." Metabolomics Workbench : Home, January 30, 2020. https://www.metabolomicsworkbench.org/.

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

———. "BLAST Topics." National Center for Biotechnology Information is part of the United States National Library of Medicine, a branch of the National Institutes of Health, 2020. https://fairsharing.org/FAIRsharing.rz4vfg. https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=BlastHelp.

———. "Database of Genotypes and Phenotypes (DbGaP)." National Center for Biotechnology Information is part of the United States National Library of Medicine, a branch of the National Institutes of Health, 2020. https://fairsharing.org/FAIRsharing.88v2k0. https://www.ncbi.nlm.nih.gov/gap/.

———. "GenBank." National Center for Biotechnology Information is part of the United States National Library of Medicine, a branch of the National Institutes of Health, 2013. https://fairsharing.org/FAIRsharing.9kahy4. https://www.ncbi.nlm.nih.gov/genbank/.

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

———. "RefSeq: NCBI Reference Sequence Database." RefSeq: NCBI Reference Sequence Database, 2017. https://www.ncbi.nlm.nih.gov/refseq/.

———. "Sequence Read Archive (SRA)." National Center for Biotechnology Information is part of the United States National Library of Medicine, a branch of the National Institutes of Health, October 3, 2019. https://fairsharing.org/FAIRsharing.g7t2hv. https://www.ncbi.nlm.nih.gov/sra/.

———. *The NCBI Handbook*. 2nd ed. National Center for Biotechnology Information (US), 2013.

———. "Viral Genomes." National Center for Biotechnology Information is part of the United States National Library of Medicine, a branch of the National Institutes of Health, 2020. https://fairsharing.org/FAIRsharing.qt5ky7. https://www.ncbi.nlm.nih.gov/genome/viruses/.

Nelson, C., N. Lurie, J. Wasserman, and S. Zakowski. "Conceptualizing and Defining Public Health Emergency Preparedness." *American Journal of Public Health* 97 Suppl 1 (2007): S9-11. https://doi.org/10.2105/AJPH.2007.114496.

nestor. "Nestor - Seal for Trustworthy Digital Archives," 2020. https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html.

Ng, Victoria, and Jan M. Sargeant. "A Quantitative Approach to the Prioritization of Zoonotic Diseases in North America: A Health Professionals' Perspective." *PLOS ONE*. 1160 BATTERY STREET, STE 100, SAN FRANCISCO, CA 94111 USA: PUBLIC LIBRARY SCIENCE, August 21, 2013. https://doi.org/10.1371/journal.pone.0072172.

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.

Nickerson, David, Koray Atalag, Bernard de Bono, Jörg Geiger, Carole Goble, Susanne Hollmann, Joachim Lonien, et al. "The Human Physiome: How Standards, Software and Innovative Service Infrastructures Are Providing the Building Blocks to Make It Achievable." *Interface Focus* 6, no. 2 (April 6, 2016): 20150103. https://doi.org/10.1098/rsfs.2015.0103.

Nickerson, M. "First Nation's Data Governance: Measuring the Nation-to-Nation Relationship Discussion Paper," 2017. https://static1.squarespace.com/static/558c624de4b0574c94d62a61/t/5ade9674575d1fb25a1c873b/1524536949054/NATION-TO-NATION_FN_DATA_GOVERNANCE_-_FINAL_-_EN.DOCX.

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.

———. "COVID-19 OBSSR Research Tools." National Institutes of Health, May 14, 2020. https://www.nlm.nih.gov/dr2/COVID-19_BSSR_Research_Tools.pdf.

———. "NIH Public Health Emergency and Disaster Research Response (DR2) COVID-19 Research Tools - Training Material." NIH Public Health Emergency and Disaster Research Response (DR2), 2020. https://dr2.nlm.nih.gov/.

———. "NIH to Host Webinar on Sharing, Discovering, and Citing COVID-19 Data and Code in Generalist Repositories on April 24 | Data Science at NIH," April 30, 2020. https://datascience.nih.gov/news/nih-to-host-webinar-on-sharing-discovering-and-citing-covid-19-data-and-code-in-generalist-repositories-on-april-24.

———. "NOT-OD-20-073: Notice of Special Interest (NOSI): Administrative Supplements to Support Enhancement of Software Tools for Open Science," April 30, 2020. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-073.html.

———. "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/.

Nii-Trebi, Nicholas Israel. "Emerging and Neglected Infectious Diseases: Insights, Advances, and Challenges." *BioMed Research International* 2017 (2017). https://doi.org/10.1155/2017/5245021.

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.

NIST PWG. "Big Data Interoperability Framework: Volume 1, Definitions." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-1r2.

———. "Big Data Interoperability Framework: Volume 2, Big Data Taxonomies." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, November 2019. https://doi.org/10.6028/NIST.SP.1500-2r2.

———. "Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-3r2.

———. "Big Data Interoperability Framework: Volume 4, Security and Privacy." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-4r2.

———. "Big Data Interoperability Framework: Volume 5, Architectures White Paper Survey." National Institute of Standards and Technology, Big Data Public Working Group, October 2015. https://doi.org/10.6028/NIST.SP.1500-5.

———. "Big Data Interoperability Framework: Volume 6, Reference Architecture." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-6r2.

———. "Big Data Interoperability Framework: Volume 7, Standards Roadmap." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-7r2.

———. "Big Data Interoperability Framework: Volume 8, Reference Architecture Interfaces." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-9r1.

———. "Big Data Interoperability Framework: Volume 9, Adoption and Modernization." Gaithersburg, MD: National Institute of Standards and Technology, Big Data Public Working Group, October 2019. https://doi.org/10.6028/NIST.SP.1500-10r1.

NMReDATA initiative. "NMReDATA." Scope | NMReDATA initiative, June 16, 2017. http://nmredata.org/.

NSW Health. "Novel-Coronavirus-Case-Questionnaire.Pdf," March 23, 2020. https://www.health.nsw.gov.au/Infectious/Forms/novel-coronavirus-case-questionnaire.pdf.

Nüst, Daniel, Vanessa Sochat, Ben Marwick, Stephen Eglen, Tim Head, Tony Hirst, and Benjamin Evans. "Ten Simple Rules for Writing Dockerfiles for Reproducible Data Science." Preprint, April 17, 2020. Open Science Framework DOI: 10.31219/osf.io/fsd7t. https://osf.io/fsd7t.

NYC Health. *Nychealth/Coronavirus-Data*. 2020. Reprint, NYC Department of Health and Mental Hygiene, 2020. https://github.com/nychealth/coronavirus-data.

NYT. "Coronavirus (Covid-19) Data in the United States." New York Times, April 21, 2020. https://github.com/nytimes/covid-19-data.

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/.

———. "Recommendation of the Council on Health Data Governance," 2019. https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf.

Office of the High Commissioner for Human Rights. "OHCHR | Annual Reports," 2020. https://www.ohchr.org/EN/Issues/Privacy/SR/Pages/AnnualReports.aspx.

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.

OGP. "Policy Areas." Open Government Partnership, 2020. https://www.opengovpartnership.org/policy-areas/.

———. "Statement on the COVID-19 Response from Civil Society Members of OGP Steering Committee." Open Government Partnership, 2020. https://www.opengovpartnership.org/news/statement-on-the-covid-19-response-from-civil-society-members-of-ogp-steering-committee/.

Ogunyemi, Omolola I., Daniella Meeker, Hyeon-Eui Kim, Naveen Ashish, Seena Farzaneh, and Aziz Boxwala. "Identifying Appropriate Reference Data Models for Comparative Effectiveness Research (CER) Studies Based on Data from Clinical Information Systems." *MEDICAL CARE* 51, no. 8, 3 (August 2013): S45–52. https://doi.org/10.1097/MLR.0b013e31829b1e0b.

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, Rita Banzi, Steve Canham, Serena Battaglia, Mihaela Matei, Christopher Ariyo, Lauren Becnel, et al. "Sharing and Reuse of Individual Participant Data from Clinical Trials: Principles and Recommendations." *BMJ Open* 7, no. 12 (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.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books, 2016. https://dl.acm.org/doi/book/10.5555/3175762.

Open ICPSR. "COVID-19 Data Repository," 2020. https://www.openicpsr.org/openicpsr/covid19.

———. "OpenICPSR: Share Your Behavioral Health and Social Science Research Data," 2020. https://www.openicpsr.org/openicpsr/.

OpenAIRE. "ARGOS Data Management Plans Creator," 2015. https://argos.openaire.eu/home.

———. "COVID-19 Open Research Gateway." OpenAIRE - Connect, 2020. https://beta.covid-19.openaire.eu/content.

OPIDoR. "DMP OPIDoR," 2016. https://dmp.opidor.fr/.

Organisation for Economic Co-operation and Development. "Why Open Science Is Critical to Combatting COVID-19 - OECD," May 12, 2020. https://read.oecd-ilibrary.org/view/?ref=129_129916-31pgjnl6cb&title=Why-open-science-is-critical-to-combatting-COVID-19.

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.

Panchadsaram, Ryan, Natalie Davis, Kristin Wikelius, Cyrus Shahpar, Laura Cobb, Marta Wosińska, David Anderson, Lucas Merrill Brown, Devin Hunt, and David Goligorsky. "COVID Exit Strategy: How We Reopen Safely," 2020. https://www.covidexitstrategy.org/.

Park, Han Woo, Sejung Park, and Miyoung Chong. "Conversations and Medical News Frames on Twitter: Infodemiological Study on COVID-19 in South Korea." *JOURNAL OF MEDICAL INTERNET RESEARCH* 22, no. 5 (May 5, 2020). https://doi.org/10.2196/18897.

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 (January 8, 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.

Pedraza, Pablo de, and Ian Vollbracht. "The Semicircular Flow of the Data Economy." *Publications Office of the European Union*, 2019, 47. https://doi.org/doi:10.2760/668.

Pedro-Roig, Laia, and Christoph H. Emmerich. "The Reproducibility Crisis in Preclinical Research – Lessons to Learn from Clinical Research." *Medical Writing* 26 (2017): 28–32.

Peirlinck, Mathias, Kevin Linka, Francisco Sahli Costabal, and Ellen Kuhl. "Outbreak Dynamics of COVID-19 in China and the United States." *BIOMECHANICS AND MODELING IN MECHANOBIOLOGY*, April 27, 2020. https://doi.org/10.1007/s10237-020-01332-5.

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.

periCOVID Uganda CRF. "PeriCOVID Uganda CRF.Docx." Dropbox, April 10, 2020. https://www.dropbox.com/s/1p32oudodv8bm1h/periCOVID%20Uganda%20CRF.docx?dl=0.

Persons, Kenneth R., Jason Nagels, Chris Carr, David S. Mendelson, Henri Rik Primo, Bernd Fischer, and Matthew Doyle. "Interoperability and Considerations for Standards-Based Exchange of Medical

Images: HIMSS-SIIM Collaborative White Paper." *Journal of Digital Imaging* 33, no. 1 (February 2020): 6–16. https://doi.org/10.1007/s10278-019-00294-0.

PhenoMeNal H2020 project. "NmrML." nmrML - home, 2019. http://nmrml.org/.

PhenoMeNal H2020 project, and COSMOS FP7- COordination Of Standards In MetabOlomicS Project. *NmrML Schema and NMR Ontology* (version v1.0.rc1). HTML. 2012. Reprint, nmrML, 2019. https://github.com/nmrML/nmrML.

PhenX. "PhenX Toolkit - COVID-19 Protocols," 2020. https://www.phenxtoolkit.org/covid19.

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.

Plowright, Raina K., Colin R. Parrish, Hamish McCallum, Peter J. Hudson, Albert I. Ko, Andrea L. Graham, and James O. Lloyd-Smith. "Pathways to Zoonotic Spillover." *Nature Reviews Microbiology* 15, no. 8 (August 2017): 502–10. https://doi.org/10.1038/nrmicro.2017.45.

Portage. "DMP Assistant," 2019. https://assistant.portagenetwork.ca/en.

Priyadarsini, S. Lakshmi, and M. Suresh. "Factors Influencing the Epidemiological Characteristics of Pandemic COVID 19: A TISM Approach." *INTERNATIONAL JOURNAL OF HEALTHCARE MANAGEMENT*, April 20, 2020. https://doi.org/10.1080/20479700.2020.1755804.

Pruitt, Kim D., Tatiana Tatusova, Garth R. Brown, and Donna R. Maglott. "NCBI Reference Sequences (RefSeq): Current Status, New Features and Genome Annotation Policy." *Nucleic Acids Research* 40, no. Database issue (January 2012): D130-135. https://doi.org/10.1093/nar/gkr1079.

PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd. "ISO 16363." PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd, 2014. 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.

Raimondi, Manuela T., Francesca Donnaloja, Bianca Barzaghini, Alberto Bocconi, Claudio Conci, Valentina Parodi, Emauela Jacchetti, et al. "Bioengineering Tools to Speed up the Discovery and Preclinical Testing of Vaccines for SARS-CoV-2 and Therapeutic Agents for COVID-19." *Theranostics* 10, no. 15 (2020): 7034–52.

Rainie, Stephanie Carroll, Jennifer Lee Schultz, Eileen Briggs, Patricia Riggs, and Nancy Lynn Palmanteer-Holder. "Data as a Strategic Resource: Self-Determination, Governance, and the Data Challenge for Indigenous Nations in the United States." *International Indigenous Policy Journal* 8, no. 2 (March 10, 2017). https://doi.org/10.18584/iipj.2017.8.2.1.

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

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

Ray, Debashree, Maxwell Salvatore, Rupam Bhattacharyya, Lili Wang, Jiacong Du, Shariq Mohammed, Soumik Purkayastha, et al. "Predictions, Role of Interventions and Effects of a Historic National Lockdown in India's Response to the the COVID-19 Pandemic: Data Science Call to Arms." *Harvard Data Science Review*, 2020. https://doi.org/10.1162/99608f92.60e08ed5.

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. "CARE Principles of Indigenous Data Governance," November 18, 2018. https://www.gida-global.org/care.

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 Zotero WG. "RDA-COVID19 Zotero Library." Research Data Alliance, 2020. https://doi.org/10.15497/rda00051.

RDA-COVID19-Epidemiology WG. "Data Sharing in Epidemiology (Version 0.053)." Research Data Alliance, 2020. https://doi.org/10.15497/rda00049.

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

RDA-COVID19-WG. "Recommendations and Guidelines." Research Data Alliance, 2020. https://doi.org/10.15497/rda00052.

Refugees, United Nations High Commissioner for. "United Nations Declaration on the Rights of Indigenous Peoples," 2007. https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html.

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 (RDA). "Get Involved | RDA," March 22, 2016. https://www.rd-alliance.org/get-involved.html.

Ricciardi, Francesca, Paola De Bernardi, and Valter Cantino. "System Dynamics Modeling as a Circular Process: The Smart Commons Approach to Impact Management." *Technological Forecasting and Social Change* 151, no. C (2020). https://ideas.repec.org/a/eee/tefoso/v151y2020ics0040162519310923.html.

Rist, Cassidy Logan, Carmen Sofia Arriola, and Carol Rubin. "Prioritizing Zoonoses: A Proposed One Health Tool for Collaborative Decision-Making." *PLOS ONE*. 1160 BATTERY STREET, STE 100, SAN FRANCISCO, CA 94111 USA: PUBLIC LIBRARY SCIENCE, October 10, 2014. https://doi.org/10.1371/journal.pone.0109986.

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.

Rockefeller Foundation. "Call for Entries: Data Science Breakthroughs for an Inclusive Recovery." The Rockefeller Foundation, May 19, 2020. https://www.rockefellerfoundation.org/blog/call-for-entries-data-science-breakthroughs-for-an-inclusive-recovery/.

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.

Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. "Ten Simple Rules for Reproducible Computational Research." Edited by Philip E. Bourne. *PLoS Computational Biology* 9, no. 10 (October 24, 2013): e1003285. https://doi.org/10.1371/journal.pcbi.1003285.

Sansone, Susanna-Assunta, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, and Orlaith Burke. "STATO: An Ontology of Statistical Methods," 2018. https://fairsharing.org/FAIRsharing.na5xp. http://stato-ontology.org/.

Sansone, Susanna-Assunta, Peter McQuilton, Philippe Rocca-Serra, Alejandra Gonzalez-Beltran, Massimiliano Izzo, Allyson L. Lister, and Milo Thurston. "FAIRsharing as a Community Approach to Standards, Repositories and Policies." *Nature Biotechnology* 37, no. 4 (April 2019): 358–67. https://doi.org/10.1038/s41587-019-0080-8.

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

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.

Sauermann, Stefan, Chifundo Kanjala, Matthias Templ, and the RDA-COVID19-WG. "Preservation of Individuals' Privacy in Shared COVID-19 Related Data." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

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.

schema.org. "Home - Schema.Org," May 1, 2020. https://schema.org/.

Schmidt, Carsten O, Rajini Nagrani, Christina Stange, Matthias Löbe, Atinkut Zeleke, Guillaume Fabre, Sofiya Koleva, et al. "COVID-19 Questionnaires, Surveys and Item-Banks: Overview of Clinical- and Population-Based Instruments." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

Schroeder, Doris. "A Global Ethics Code to Fight 'ethics Dumping' in Research," 2020. https://www.globalcodeofconduct.org/.

SDMX initiative. "Standards | SDMX – Statistical Data and Metadata EXchange," March 2, 2020. https://sdmx.org/?page_id=5008.

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

Setti, Leonardo, Fabrizio Passarini, Gianluigi De Gennaro, Pierluigi Baribieri, Maria Grazia Perrone, Massimo Borelli, Jolanda Palmisani, et al. "SARS-Cov-2 RNA Found on Particulate Matter of Bergamo in Northern Italy: First Preliminary Evidence." Cold Spring Harbor Laboratory Press, April 24, 2020. https://www.medrxiv.org/content/10.1101/2020.04.15.20065995v2.

Shafranovich <ietf@shaftek.org>, Yakov. "Common Format and MIME Type for Comma-Separated Values (CSV) Files," October 2005. https://tools.ietf.org/html/rfc4180.

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.

Sharing Rewards and Credit Interest Group, Research Data Alliance. "Sharing Rewards and Credit (SHARC) IG." RDA, August 7, 2017. https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig.

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.

Shiferaw, Miriam L., Jeffrey B. Doty, Giorgi Maghlakelidze, Juliette Morgan, Ekaterine Khmaladze, Otar Parkadze, Marina Donduashvili, et al. "Frameworks for Preventing, Detecting, and Controlling Zoonotic Diseases - Volume 23, Supplement—December 2017 - Emerging Infectious Diseases Journal - CDC," 2017. https://doi.org/10.3201/eid2313.170601.

Skyline. "Panorama: Repository Software for Targeted Mass Spectrometry Assays from Skyline," 2018. https://panoramaweb.org/project/home/begin.view.

Smiley Evans, T., Z. Shi, M. Boots, W. Liu, K.J. Olival, X. Xiao, S. Vandewoude, et al. "Synergistic China–US Ecological Research Is Essential for Global Emerging Infectious Disease Preparedness." *EcoHealth* 17, no. 1 (2020): 160–73. https://doi.org/10.1007/s10393-020-01471-2.

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

Smith, K.F., M. Behrens, L.M. Schloegel, N. Marano, S. Burgiel, and P. Daszak. "Reducing the Risks of the Wildlife Trade." *Science* 324, no. 5927 (2009): 594–95. https://doi.org/10.1126/science.1174460.

SNOMED. "COVID-19 Data Coding Using SNOMED CT - COVID-19 Guide - SNOMED Confluence," 2020. https://confluence.ihtsdotools.org/display/DOCCV19.

———. "SNOMED Home Page." SNOMED, 2020. /.

Spaaks, Jurriaan H., Benjamin Uekermann, and Cunliang Geng. "NLeSC/Awesome-Research-Software-Registries." 2019. Reprint, Netherlands eScience Center, March 31, 2020. https://github.com/NLeSC/awesome-research-software-registries.

Spidlen, Josef, Ryan Brinkman, and ISAC Data Standards Task Force. "Gating-ML 2.0," March 16, 2015. https://fairsharing.org/FAIRsharing.qpyp5g. http://flowcyt.sourceforge.net/gating/.

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.

Staunton, Ciara, Santa Slokenberga, and Deborah Mascalzoni. "The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks." *European Journal of Human Genetics* 27, no. 8 (August 2019): 1159–67. https://doi.org/10.1038/s41431-019-0386-5.

Stein, Lincoln. "GFF and GVF Specification Documents," February 26, 2013. https://fairsharing.org/FAIRsharing.dnk0f6. https://github.com/The-Sequence-Ontology/Specifications.

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.

Sun, Pengfei, Xiaosheng Lu, Chao Xu, Wenjuan Sun, and Bo Pan. "Understanding of COVID-19 Based on Current Evidence." *Journal of Medical Virology* 92, no. 6 (2020): 548–51. https://doi.org/10.1002/jmv.25722.

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.

Task Force on, and Privacy and the Protection of Health-Related Data. "UNSRPhealthrelateddataRecCLEAN.Pdf," December 5, 2019.

https://www.ohchr.org/Documents/Issues/Privacy/SR_Privacy/UNSRPhealthrelateddataRecCLEAN.pdf.

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.

Taylor, L.H., S.M. Latham, and M.E.J. Woolhouse. "Risk Factors for Human Disease Emergence." *Philosophical Transactions of the Royal Society B: Biological Sciences* 356, no. 1411 (2001): 983–89. https://doi.org/10.1098/rstb.2001.0888.

Taylor, Linnet, Luciano Floridi, and Bart van der Sloot. *Group Privacy: New Challenges of Data Technologies*. Philosophical Studies Series. Springer, 2017. https://link.springer.com/book/10.1007%2F978-3-319-46608-8.

Taylor, Mark T. *Genetic Data and the Law: A Critical Perspective on Privacy Protection*. Cambridge Bioethics and Law (16). CAMBRIDGE UNIV PRESS, THE PITT BUILDING, TRUMPINGTON ST, CAMBRIDGE CB2 1RP, CAMBS, ENGLAND, 2012. https://doi.org/10.1017/CBO9780511910128.

Te Pūnaha Matatini. "COVID-19 Infection Fatality Rates by Ethnicity for New Zealand," April 17, 2020. https://www.tepunahamatatini.ac.nz/2020/04/17/estimated-inequities-in-covid-19-infection-fatality-rates-by-ethnicity-for-aotearoa-new-zealand/.

Te Rōpū Whakakaupapa Urutā. "MR_Māori Response Action Plan." Urutā, April 20, 2020. https://www.uruta.maori.nz/maori-response-action-plan.

Team, GESIS Panel. *GESIS Panel Special Survey on the Coronavirus SARS-CoV-2 Outbreak in Germany*, 2020. https://doi.org/10.4232/1.13520.

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, M. "Quality Indicators for Statistical Disclosure Methods: A Case Study on the Structure of Earnings Survey." *Journal of Official Statistics* 31, no. 4 (2015): 737–61. https://doi.org/10.1515/JOS-2015-0043.

———. *Statistical Disclosure Control for Microdata: Methods and Applications in R*. Statistical Disclosure Control for Microdata: Methods and Applications in R. Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-50272-4.

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.

Templ, Matthias, Bernhard Meindl, Alexander Kowarik, and Shuang Chen. "Introduction to Statistical Disclosure Control (SDC)," IHSN Working Paper No 007, 2014, 25.

The Adaptive Immune Receptor Repertoire Community. "The AIRR Community." The Antibody Society, March 26, 2020. https://www.antibodysociety.org/the-airr-community/.

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

———. "The COVID Racial Data Tracker." The COVID Tracking Project, 2020. https://covidtracking.com/race.

The Carpentries. "The Carpentries." The Carpentries, 2020. https://carpentries.org/index.html.

The European Data Protection Board. "Guidelines in the Processing of Data Concerning Health for the Purpose of Scientific Research in the Context of the COVID-19 Outbreak," April 21, 2020. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificr esearchcovid19_en.pdf.

The European Health Data & Evidence Network. "COVID-19-Data-Partner-Call-Description-v1.4.Pdf," April 15, 2020. https://www.ehden.eu/wp-content/uploads/2020/04/COVID-19-Data-Partner-Call-Description-v1.4.pdf.

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 National Institute of Environmental Health Science. "COVID-19_BSSR_Research_Tools.Pdf," May 14, 2020. https://www.nlm.nih.gov/dr2/COVID-19_BSSR_Research_Tools.pdf.

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.

Thorogood, Adrian, and Michael Beauvais. "Responsible Data Sharing to Respond to the COVID-19 Pandemic: Ethical and Legal Considerations (Document under Development)," 2020. https://docs.google.com/document/d/1wK_NoNYXKy0ttTQ-ySHh3ZRpvPrLV4uPwV8FSq6BQ60/edit?usp=embed_facebook.

Tim F. Rayner et al. "FGED: MAGE-TAB," 2006. http://fged.org/projects/mage-tab/.

Tonnang, E. Z., Jay Greenfield, Gary Mazzaferro, Claire C Austin, and the RDA-COVID19-WG. "COVID-19 Emergency Public Health and Economic Measures Causal Loops: A Computable Framework." In *COVID-19 Data Sharing in Epidemiology, Version 0.053*. Research Data Alliance RDA-COVID19-Epidemiology WG, 2020. https://doi.org/10.15497/rda00049.

Toronto International Data Release Workshop Authors, Ewan Birney, Thomas J. Hudson, Eric D. Green, Chris Gunter, Sean Eddy, Jane Rogers, et al. "Prepublication Data Sharing." *Nature* 461, no. 7261 (2009): 168–70. https://doi.org/10.1038/461168a.

Tscharntke, Teja, Michael E Hochberg, Tatyana A Rand, Vincent H Resh, and Jochen Krauss. "Author Sequence and Credit for Contributions in Multiauthored Publications." *PLoS Biology* 5, no. 1 (January 2007). https://doi.org/10.1371/journal.pbio.0050018.

UCSC. "Browser Extensible Data Format (BED)," 2018 2000. https://fairsharing.org/FAIRsharing.mwmbpq. http://genome.ucsc.edu/FAQ/FAQformat.html#format1.

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

UK Data Archive. "Managing Data," 2020. https://www.data-archive.ac.uk/managing-data/.

UK Data Service. "Recommended Formats," 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.

———. "UK Data Service Elsst." UK Data Service Elsst, 2019. https://elsst.ukdataservice.ac.uk/.

———. "UK Data Service Hasset." UK Data Service Hasset, 2019. https://hasset.ukdataservice.ac.uk/.

UK Research and Innovation. "Get Funding for Ideas That Address COVID-19," April 30, 2020. https://www.ukri.org/funding/funding-opportunities/ukri-open-call-for-research-and-innovation-ideas-to-address-covid-19/.

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

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

———. "Recommendations on Data and Indicators | United Nations For Indigenous Peoples." United Nations Department of Economic and Social Affairs, 2017. https://www.un.org/development/desa/indigenouspeoples/mandated-areas1/data-and-indicators/recs-data.html.

———. "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/.

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

UNESCO. "Report of the IBC on the Principle of the Sharing of Benefits." UNESCO International Bioethics Committee, October 15, 2015. https://unesdoc.unesco.org/ark:/48223/pf0000233230.

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

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

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=*.

United Nations. "Coronavirus Impact on World's Indigenous, Goes Well beyond Health Threat." UN News, May 18, 2020. https://news.un.org/en/story/2020/05/1064322.

Universidade Federal de Pelotas. "Brazil COVID Serological Survey Questionnaire 20200428.Docx." Dropbox, 2020. https://www.dropbox.com/s/l9hhblb83ybr70u/Brazil%20COVID%20serological%20survey%20questionnaire%2020200428.docx?dl=0.

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

———. "DMPTool," 2019. https://cdlib.org/services/uc3/dmptool/.

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

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

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." Data repository, 2020. https://github.com/beoutbreakprepared.

University of Washington - HGIS Lab. "Novel Coronavirus Infection Map." University of Washingtion - Humanistic GIS Laboratory, 2020. https://github.com/jakobzhao/virus.

U.S. Department of Health and Human Services. "Pan-Flu-Report-2017v2.Pdf," December 2017. https://www.cdc.gov/flu/pandemic-resources/pdf/pan-flu-report-2017v2.pdf.

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)," April 23, 2020. https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/PGD.

Vitae. "Concordat to Support the Career Development of Researchers." Landing page, April 30, 2020. https://www.vitae.ac.uk/policy/concordat-to-support-the-career-development-of-researchers.

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

Vollmer, Nicholas. "Article 9 EU General Data Protection Regulation (EU-GDPR)." Text. SecureDataService, September 5, 2018. https://www.privacy-regulation.eu/en/article-9-processing-of-special-categories-of-personal-data-GDPR.htm.

Voss, Erica A., Rupa Makadia, Amy Matcho, Qianli Ma, Chris Knoll, Martijn Schuemie, Frank J. DeFalco, Ajit Londhe, Vivienne Zhu, and Patrick B. Ryan. "Feasibility and Utility of Applications of the Common Data Model to Multiple, Disparate Observational Health Databases." *JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION* 22, no. 3 (May 2015): 553–64. https://doi.org/10.1093/jamia/ocu023.

Waagmeester, Andra, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi L Griffith, Kristina Hanspers, et al. "Wikidata as a Knowledge Graph for the Life Sciences." Edited by Peter Rodgers and Chris Mungall. *ELife* 9 (March 17, 2020): e52614. https://doi.org/10.7554/eLife.52614.

Wang, Lucy Lu, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, et al. "CORD-19: The Covid-19 Open Research Dataset." *ArXiv:2004.10706 [Cs]*, April 24, 2020. http://arxiv.org/abs/2004.10706.

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.

Warren, Matthew. "CDISC Interim User Guide for COVID-19 - CDISC Interim User Guide for COVID-19 - Wiki," April 20, 2020. https://wiki.cdisc.org/display/COVID19/CDISC+Interim+User+Guide+for+COVID-19.

Webster, Robert G. "Wet Markets - A Continuing Source of Severe Acute Respiratory Syndrome and Influenza?" *The Lancet* 363, no. 9404 (January 17, 2004): 234–36. https://doi.org/10.1016/S0140-6736(03)15329-9.

Weizmann Institute. "COVID-19: Daily Reports (דיווח קורונה יומי)," 2020. https://coronaisrael.org/en/.

Wellcome. "Final UK Covid Questionnaire_23 April.Pdf." Dropbox, April 23, 2020. https://www.dropbox.com/s/hy6jdpgvkgpgxfi/Final%20UK%20Covid%20Questionnaire_23%20April.pdf?dl=0.

Wellcome Trust. "Data, Software and Materials Management and Sharing Policy," 2017. https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy.

———. "Longitudinal Population Studies Strategy," 2017. https://wellcome.ac.uk/sites/default/files/longitudinal-population-studies-strategy_0.pdf.

WHO. "About EPI-WIN," 2020. https://www.who.int/teams/risk-communication/about-epi-win.

———. "Climate Change and Infectious Diseases." Climate change and human health. World Health Organization, 2003. https://www.who.int/globalchange/summary/en/index5.html.

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

———. "COVID-19 Global Literature on Coronavirus Disease." World Health Organization, 2019. https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/.

———. "COVID-19 Situation Reports." Dataset, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

———. "C-TAP: COVID-19 Technology Access Pool," May 29, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/covid-19-technology-access-pool.

———. "Developing Global Norms for Sharing Data and Results during Public Health Emergencies." World Health Organization. World Health Organization, September 2, 2015. http://www.who.int/medicines/ebola-treatment/data-sharing_phe/en/.

———. "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.

———. "Ethical Considerations to Guide the Use of Digital Proximity Tracking Technologies for COVID-19 Contact Tracing," 2020. https://www.who.int/publications-detail/WHO-2019-nCoV-Ethics_Contact_tracing_apps-2020.1WHO.

———. "Global Early Warning System for Major Animal Diseases, Including Zoonoses (GLEWS)." World Health Organization, 2006. https://www.who.int/foodsafety/areas_work/zoonose/glews/en/.

———. "Global Surveillance for COVID-19 Caused by Human Infection with COVID-19 Virus: Interim Guidance," 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.

———. "One Health," 2017. https://www.who.int/news-room/q-a-detail/one-health.

———. "Population-Based Age-Stratified Seroepidemiological Investigation Protocol for COVID-19 Virus Infection," 2020. https://apps.who.int/iris/bitstream/handle/10665/332188/WHO-2019-nCoV-Seroepidemiology-2020.2-eng.pdf?sequence=1&isAllowed=y.

———. "Preparing GISRS for the Upcoming Influenza Seasons during the COVID-19 Pandemic – Practical Considerations," May 26, 2020. https://apps.who.int/iris/bitstream/handle/10665/332198/WHO-2019-nCoV-Preparing_GISRS-2020.1-eng.pdf.

———. "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).

———. "Risk Communication: EPI-WIN," 2020. https://www.who.int/teams/team-preview/risk-communciation---global.

———. "Surveillance, Rapid Response Teams, and Case Investigation." Coronavirus disease (COVID-19) technical guidance, March 20, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/surveillance-and-case-definitions.

———. "Survey Tool and Guidance: Behavioural Insights on COVID-19." WHO Regional Office for Europe, 2020. http://www.euro.who.int/__data/assets/pdf_file/0007/436705/COVID-19-survey-tool-and-guidance.pdf?ua=1.

———. "WHO | Global Influenza Surveillance and Response System (GISRS)." WHO. World Health Organization, 2020. http://www.who.int/influenza/gisrs_laboratory/en/.

———. "WHO | International Classification of Diseases, 11th Revision (ICD-11)." World Health Organization. World Health Organization, 2019. http://www.who.int/classifications/icd/en/.

———. "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.

———. "Zoonoses." WHO. World Health Organization, July 19, 2017. http://www.who.int/topics/zoonoses/en/.

WHO, FAO and OIE. "Taking a Multisectoral, One Health Approach: A Tripartite Guide to Addressing Zoonotic Diseases in Countries," 2019. https://www.oie.int/fileadmin/Home/eng/Media_Center/docs/EN_TripartiteZoonosesGuide_webversion.pdf.

Wicher, Daniel. "The Covid-19 Case as an Example of Systems Thinking Usage," 2020. https://agilejar.com/2020/03/the-covid-19-case-as-an-example-of-systems-thinking-usage/.

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. https://doi.org/10.1007/978-1-4613-0121-9.

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. "Good Enough Practices in Scientific Computing." *PLOS Computational Biology* 13, no. 6 (2017): e1005510. https://doi.org/10.1371/journal.pcbi.1005510.

Wingara, Maiam Nayri. "Indigenous Data Sovereignty. Data for Governance: Governance of Data." Australian Indigenous Governance Institute, 2018.

Woo, Patrick CY, Susanna KP Lau, and Kwok-yung Yuen. "Infectious Diseases Emerging from Chinese Wet-Markets: Zoonotic Origins of Severe Respiratory Viral Infections." *Current Opinion in Infectious Diseases* 19, no. 5 (October 2006): 401–407. https://doi.org/10.1097/01.qco.0000244043.08264.fc.

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

World Wide Web Consortium. "Data Catalog Vocabulary (DCAT) - Version 2," February 4, 2020. https://www.w3.org/TR/vocab-dcat-2/.

Worldometer. "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/.

Wu, D., T. Wu, Q. Liu, and Z. Yang. "The SARS-CoV-2 Outbreak: What We Know." *International Journal of Infectious Diseases* 94 (2020): 44–48. https://doi.org/10.1016/j.ijid.2020.03.004.

RDA COVID-19 Recommendations and Guidelines on Data Sharing, 30 June 2020

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.

Yang, Chenglei, Xue Qiu, Haoran Fan, Mei Jiang, Xiaojie Lao, Yukeng Zeng, and Zhiming Zhang. "Coronavirus Disease 2019: Reassembly Attack of Coronavirus." *INTERNATIONAL JOURNAL OF ENVIRONMENTAL HEALTH RESEARCH*, April 21, 2020. https://doi.org/10.1080/09603123.2020.1747602.

Yang, Tong, Kai Shen, Sixuan He, Enyu Li, Peter Sun, Pingying Chen, Lin Zuo, et al. "CovidNet: 1Point3Acres," 2020. https://coronavirus.1point3acres.com/en.

———. "CovidNet: To Bring Data Transparency in the Era of COVID-19," June 4, 2020. http://arxiv.org/abs/2005.10948.

Yasaka, Tyler M., Brandon M. Lehrich, and Ronald Sahyouni. "Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App." *JMIR MHEALTH AND UHEALTH* 8, no. 4 (April 7, 2020). https://doi.org/10.2196/18936.

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.

Zastrow, Mark. "Open Science Takes on the Coronavirus Pandemic." *Nature* 581, no. 7806 (April 24, 2020): 109–10. https://doi.org/10.1038/d41586-020-01246-3.

Zhang, Alison. *Creating Digital Collections: A Practical Guide - 1st Edition*, 2008. https://www.elsevier.com/books/creating-digital-collections/zhang/978-1-84334-396-7.

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, Lei, Sepehr Ghader, Michael L. Pack, Chenfeng Xiong, Aref Darzi, Mofeng Yang, Qianqian Sun, AliAkbar Kabiri, and Songhua Hu. "An Interactive COVID-19 Mobility Impact and Social Distancing Analysis Platform." *MedRxiv*, May 5, 2020, 2020.04.29.20085472. https://doi.org/10.1101/2020.04.29.20085472.

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.

Zheng, Wei-Shi, Shaogang Gong, and Tao Xiang. "Person Re-Dentification by Probabilistic Relative Distance Comparison," 649–56. Providence, RI: IEEE, 2011. https://doi.org/10.1109/CVPR.2011.5995598.

Zhou, Peng, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, et al. "A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin." *Nature* 579, no. 7798 (March 2020): 270–73. https://doi.org/10.1038/s41586-020-2012-7.

Zhou, Shuang-Jiang, Li-Gang Zhang, Lei-Lei Wang, Zhao-Chang Guo, Jing-Qi Wang, Jin-Cheng Chen, Mei Liu, Xi Chen, and Jing-Xu Chen. "Prevalence and Socio-Demographic Correlates of Psychological

Health Problems in Chinese Adolescents during the Outbreak of COVID-19." *EUROPEAN CHILD & ADOLESCENT PSYCHIATRY*, May 3, 2020. https://doi.org/10.1007/s00787-020-01541-4.

Zhou, Yiwang, Lili Wang, Leyao Zhang, Lan Shi, Kangping Yang, Jie He, Bangyao Zhao, et al. "A Spatiotemporal Epidemiological Prediction Model to Inform County-Level COVID-19 Risk in the USA." *Harvard Data Science Review*, 2020. https://hdsr.mitpress.mit.edu/pub/qqg19a0r/release/1.

15. 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 alphabetically by last name as: First-Name Last-Name ORCID

Clara Amid (0000-0001-6534-7425)
Pamela Andanda (0000-0002-2746-7861)
Claire C Austin (0000-0001-9138-5986)
Christophe Bahim
Michelle Barker (0000-0002-3623-172X)
Claudia Bauzer Medeiros (0000-0003-1908-4753)
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 (0000-0002-8900-8978)
Hélène Blasco (0000-0001-6107-0035)
Franziska Boehm
Sabrina Boni
Sergio Bonini (0000-0003-0079-3031)
Ann Borda (0000-0003-3884-2978)
Korbinian Bösl (0000-0003-0498-4273)
Christian E. Busse (0000-0001-7553-905X)
Anne Cambon-Thomsen (0000-0001-8793-3644)
Maeve Campman (0000-0002-6613-7144)
Stephanie Carroll (0000-0002-8996-8071)
Calvin Wing Yiu Chan (0000-0002-3656-7709)
Neil Chue Hong (0000-0002-8876-7606)
Pyrou Chung (0000-0002-4133-3149)
Jorge Clarke (0000-0003-1314-7020)
Gerard Coen (0000-0001-9915-9721)
Donna Cormack (0000-0003-2854-3595)
Brian Corrie (0000-0003-3888-6495)
Zoe Cournia (000-0001-9287-264X)
Andreas Czerniak (0000-0003-3883-4169)
Piotr Wojciech Dabrowski (0000-0003-4893-805X)
Pablo de Pedraza
Luc Decker (0000-0002-4808-3568)
Laurence Delhaes (0000-0001-7489-9205)
David Delmail (0000-0003-2836-6496)
Cyrille Delpierre (0000-0002-0831-080X)
Philippe Després (0000-0002-4163-7353)
Natalie Dewson (0000-0002-5968-9696)
Kheeran Dharmawardena (0000-0002-4292-7475)

Gayo Diallo (0000-0002-9799-9484)
Ingrid Dillo (0000-0001-5654-2392)
Diana Dimitrova (0000-0003-4732-7054)
Laurent Dollé (0000-0003-4566-6407)
Nora Dörrenbächer (0000-0002-6246-1051)
Stephan Druskat (0000-0003-4925-7248)
Thomas Duflot (0000-0002-8730-284X)
Patrick Dunn (0000-0003-1868-9689)
Patrice Duroux (0000-0001-8935-7900)
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 Jael Garcia Castro (0000-0003-3986-0510)
Sandra Gesing (0000-0002-6051-0673)
Veronique Giudicelli (0000-0002-2258-468X)
Carole Goble (0000-0003-1219-2137)
Martin Golebiewski (0000-0002-8683-7084)
Alejandra Gonzalez-Beltran (0000-0003-3499-8262)
Jay Greenfield (0000-0003-2773-5317)
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)
Rob Hooft (0000-0001-6825-9439)
Maui Hudson (0000-0003-3880-4015)
Wim Hugo (0000-0002-0255-5101)
Andrea Jackson Dipina
Ann Myatt James (0000-0002-2137-7961)
Sarah Jones (0000-0002-5094-7126)
Nick Jones (0000-0001-5513-8312)
Chifundo Kanjala (0000-0003-0540-8374)
Daniel S. Katz (0000-0001-5934-7525)
Sofia Kossida (0000-0003-2482-0022)
Iryna Kuchma (000-0002-2064-3439)
Tahu Kukutai (0000-0001-5080-2296)
Helena Laaksonen (0000-0002-1312-1958)
Anna-Lena Lamprecht (0000-0003-1953-5606)

Dollé Laurent (0000-0003-4566-6407)
Paula Martinez Lavanchy (0000-0003-1448-0917)
Young-Joo Lee (0000-0001-7189-6607)
Mark Leggott (0000-0003-1392-7799)
Joanna Leng (0000-0001-9790-162X)
Marcia Levenstein
Dawei Lin (0000-0002-5506-0030)
Birte Lindstaedt (0000-0002-8251-1597)
Aliaksandra Lisouskaya (0000-0001-7556-8977)
Nicolas Loozen
Tovani-Palone Marcos Roberto (0000-0003-1149-2437)
Paula Andrea Martinez (0000-0002-8990-1985)
Gary Mazzaferro (0000-0002-3196-7201)
Katherine McNeill (0000-0003-2865-3751)
Peter McQuilton (0000-0003-2687-1982)
Eva Méndez (0000-0002-5337-4722)
Natalie Meyers (0000-0001-6441-6716)
Robin Michelet (0000-0002-5485-607X)
Daniel Mietchen (0000-0001-9488-1870)
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)
Amy Nurnberger (0000-0002-5931-072X)
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 Luis Parra-Calderon (0000-0003-2609-575X)
Pandelis Perakakis (0000-0002-9130-3247)
Brian Pickering (0000-0002-6815-2938)
Amy Pienta (0000-0003-1174-6118)
Priyanka Pillai (0000-0002-3768-8895)
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)
Valeria Quochi (0000-0002-1321-5444)

Dana Rad (0000-0001-6754-3585)
Lane Rasberry (0000-0002-9485-6146)
Alessandra Renieri (0000-0002-0846-9220)
Stéphanie Rennes (0000-0003-1458-7773)
Artur Rocha (0000-0002-5637-1041)
Robyn Rowe (0000-0002-8028-5274)
Gavin Rozzi
Susanna-Assunta Sansone (0000-0001-5306-5690)
Rodrigo Sara
Venkata Satagopam (0000-0002-6532-5880)
Stefan Sauermann (0000-0003-0824-9989)
Henry Schaefer (0000-0002-3492-811X)
Carsten Oliver Schmidt (0000-0001-5266-9396)
Lynn M. Schriml (0000-0001-8910-9851)
Meg Sears (0000-0002-6987-1694)
Hugh Shanahan (0000-0003-1374-6015)
Lina Sitz (0000-0002-6333-4986)
Tim Smith (0000-0002-1567-7116)
Joanne Stocks (0000-0002-7800-6002)
Rainer Stotzka (0000-0003-3642-1264)
Shoaib Sufi (0000-0001-6390-2616)
Michele Suina (0000-0002-7661-2359)
Mark Taylor (0000-0003-2009-6284)
Marta Teperek (0000-0001-8520-5598)
Mogens Thomsen (0000-0002-4546-0129)
Henri Tonnang (0000-0002-9424-9186)
Marcos Roberto Tovani-Palone (0000-0003-1149-2437)
Susheel Turinici (0000-0003-2713-006X)
Yasemin Türkyilmaz-Van der Velden (0000-0003-2562-0452)
Mary Uhlmansiek (0000-0002-7949-2057)
Meghan Underwood (0000-0001-6538-9617)
Justine Vandendorpe (0000-0002-9421-8582)
Susheel Varma (0000-0003-1687-2754)
Bridget Walker
Maggie Walter (0000-0002-8028-5274)
Minglu Wang (0000-0002-0021-5605)
Yan Wang (0000-0002-6317-7546)
Galia Weidl
Anna Widyastuti (0000-0003-2149-935X)
Kara Woo (0000-0002-5125-4188)
Qian Zhang (0000-0003-1549-7358)

16. Acknowledgements

We would also like to thank those who provided substantive feedback on various draft releases of this document:

Abdelkrim Boujraf	Manuela T. Raimondi
Sandoval Carneiro	Phil Robinson
Kathryn Cassidy	David Romain
Peter Cornwell	Giorgio Rossi
Tim Dafoe	Merran Smith
Michele Loi	Robert Fraser Terry
Daniel Martins	Adalberto Val
Alain Paic	Nottingham University Hospitals NHS Trust, COVID-19 Patient and Public Task Force
Carme Plasencia	
Kim D. Pruitt	