

**RDA COVID-19**

**Recommendations and Guidelines**

***Relevant sections for publishers***

**RDA Recommendation (5th Release)**

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

Background

The RDA COVID-19 Working Group (CWG) members bring various 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. The full document can be [found here](https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0). This document contains the relevant parts for publishers.

1. General principles

Research outputs should align with the **FAIR principles**, meaning that data, software, models and other outputs should be Findable, Accessible, Interoperable and Reusable. However, there is also consensus that outputs need to be shared as quickly as possible in order to have a direct impact on the progress of the pandemic. A balance between achieving ‘perfectly’ FAIR outputs and timely sharing is necessary, with the key goal of immediate and open sharing as a driver. Researchers should also be encouraged to share what they have as-is without fear of it being insufficient, and signal that help is needed.

Researchers are encouraged to apply **generic metadata** for Covid-19 related data, as much as possible, such as for example any of the four generic and well-established metadata standards that are used widely, Dublin Core (DC), DCAT, DataCite and Schema.org.

Authors to add available **contextual documentation**, when sharing their research outputs, including documentation on software, documentation of methodologies used to define and construct data, data cleaning, data imputation, data provenance and so on.

Participants respect the ground rules for **Ethics, Privacy and sound Legal Frameworks** in Covid-19 data sharing, so that access to individual participant data and trial documents can be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse. Emergency data legislation activated during a pandemic needs to clearly outline data custodianship/ownership, publication rights and arrangements, consent models, and permissions around sharing data and exemptions.

More detailed information about these points can be found in the [full document](https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0).

### Trustworthy Data Repositories

To facilitate data quality control, timely sharing and sustained access, data should be deposited in data repositories. Whenever possible, these should be *trustworthy* data repositories (TDRs) that have been certified, subject to rigorous governance, and committed to longer-term preservation of their data holdings.

Examples of such certifications are [CoreTrustSeal](https://www.coretrustseal.org/) ([CoreTrustSeal](https://www.coretrustseal.org/)), [Nestor Seal for Trustworthy Digital Archives](https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html) ([nestor](https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html)) and [ISO 16363](http://www.iso16363.org) ([PTAB](http://www.iso16363.org/)). Repositories certified by CoreTrustSeal, a result of the RDA [Repository Audit and Certification DSA–WDS Partnership WG](https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html) are [listed here](https://www.coretrustseal.org/why-certification/certified-repositories/). The underlying community-based TRUST principles ([Lin et al., 2020](https://doi.org/10.1038/s41597-020-0486-7)) 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.

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

*Recommended disciplinary specific repositories*

**Virus genomics**

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](https://fairsharing.org/search/?q=genomics&content=biodbcore).

**Host genomics**

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

1. Gene expression datashould in general be retrieved from or deposited in the repositories listed below ([Blaxter et al., 2016](https://doi.org/10.1126/science.aaf7672)). To achieve load balancing, it is recommended to choose the respective regional repository. It should be noted that [INSDC](http://www.insdc.org) resources (i.e., [DDBJ](https://www.ddbj.nig.ac.jp/index-e.html), [ENA](https://www.ebi.ac.uk/ena) and [NCBI](https://www.ncbi.nlm.nih.gov/genbank/)) synchronise most of their data sets daily[[1]](#footnote-1).
   1. Transcriptomics of human subjects (requiring authorised access):
      1. [Database of Genotypes and Phenotypes](http://www.ncbi.nlm.nih.gov/gap) (dbGaP) ([Mailman et al., 2007](https://dx.doi.org/10.1038%2Fng1007-1181); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.88v2k0))
      2. [European Genome-Phenome Archive](https://ega-archive.org) (EGA) ([Lappalainen et al., 2015](https://doi.org/10.1038/ng.3312); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.mya1ff)). The corresponding non-sensitive metadata will be available through EBI [ArrayExpress](https://www.ebi.ac.uk/arrayexpress) ([Athar et al., 2019](https://doi.org/10.1093/nar/gky964); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.6k0kwd))
      3. [Japanese Genotype-phenotype Archive](https://www.ddbj.nig.ac.jp/jga) (JGA) ([Kodama et al., 2015](https://doi.org/10.1093/nar/gku1120); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.pwgf4p)).
   2. Transcriptomics (from cell lines/animals):
      1. [ArrayExpress](https://www.ebi.ac.uk/arrayexpress) ([Athar et al., 2019](https://doi.org/10.1093/nar/gky964); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.6k0kwd))
      2. [Gene Expression Omnibus](https://www.ncbi.nlm.nih.gov/geo/) ([Barrett et al., 2013](https://doi.org/10.1093/nar/gks1193); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.5hc8vt))
      3. [Genomic Expression Archive](https://www.ddbj.nig.ac.jp/gea) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.hESBcy)).
   3. Underlying reads can be retrieved from/will automatically deposited to the corresponding read archive:
      1. [DDBJ Sequence Read Archive](https://www.ddbj.nig.ac.jp/dra/index-e.html) (DRA) ([Kodama et al., 2012](https://doi.org/10.1093/nar/gkr854); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.k337f0)), for submission documentation see [here](https://www.ddbj.nig.ac.jp/dra/submission-e.html)
      2. [European Nucleotide Archive](https://www.ebi.ac.uk/ena/browser/home) (in F[AIRsharing](https://doi.org/10.25504/FAIRsharing.dj8nt8)), for submission documentation see [here](https://ena-docs.readthedocs.io/en/latest/)
      3. NCBI [Sequence Read Archive](https://www.ncbi.nlm.nih.gov/sra/) (SRA) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.g7t2hv)), for submission documentation see [here](https://www.ncbi.nlm.nih.gov/sra/docs/submit/).
   4. Microarray-based gene expression data:
      1. [ArrayExpress](https://www.ebi.ac.uk/arrayexpress) ([Athar et al., 2019](https://doi.org/10.1093/nar/gky964); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.6k0kwd))
      2. [Gene Expression Omnibus](https://www.ncbi.nlm.nih.gov/geo/) ([Barrett et al., 2013](https://doi.org/10.1093/nar/gks1193); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.5hc8vt))
      3. [Genomic Expression Archive](https://www.ddbj.nig.ac.jp/gea) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.hESBcy)).
   5. Data on the originating sample can be retrieved from/will automatically be deposited to the corresponding sample archive:
      1. [DDBJ BioSample](https://www.ddbj.nig.ac.jp/biosample/index-e.html) (in [FAIRsharing](https://fairsharing.org/biodbcore-001519))
      2. [EBI BioSamples](https://www.ebi.ac.uk/biosamples/) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.ewjdq6))
      3. [NCBI BioSample](https://www.ncbi.nlm.nih.gov/biosample/) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.qr6pqk)).
   6. For specialised use cases, additional domain-specific repositories might exist, a curated list of which can be found in [FAIRsharing](https://fairsharing.org/search/?q=transcriptomics&content=biodbcore). 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):
   1. [GWAS Catalog](https://www.ebi.ac.uk/gwas/) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.blUMRx))
   2. [EGA](https://doi.org/10.25504/FAIRsharing.mya1ff) ([Lappalainen et al., 2015](https://doi.org/10.1038/ng.3312); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.mya1ff))
   3. [GWAS Central](https://doi.org/10.25504/FAIRsharing.vkr57k) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.vkr57k)).
3. Adaptive Immune Receptor Repertoire Sequencing (AIRR-seq)[[2]](#footnote-2) data: It is recommended that data be deposited using [AIRR Community](http://www.airr-community.org) compliant processes and standards, in either of the following repositories.
   1. AIRR-seq specific repositories that are part of the [AIRR Data Commons](https://docs.airr-community.org/en/latest/api/adc.html), for example the [iReceptor Public Archive](https://fairsharing.org/FAIRsharing.ekdqe5) ([Corrie et al., 2018](https://doi.org/10.1111/imr.12666); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.ekdqe5)) or [VDJServer](https://vdjserver.org/) ([Christley et al., 2018](https://doi.org/10.3389/fimmu.2018.00976); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.nzdq0f)).
   2. INSDC repositories via NCBI SRA/Genbank, following the AIRR Community [recommended NCBI submission processes](https://docs.airr-community.org/en/latest/standards/data_submission.html#data-submission).

**Structural Data**

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](http://www.wwpdb.org/) ([Burley et al., 2019](https://dx.doi.org/10.1093/nar/gky949); in [FAIRsharing](https://dx.doi.org/https://doi.org/10.25504/FAIRsharing.mckkb4)); a collaborating cluster of three regional centres at (i) for Europe: EBI [PDBe](https://www.ebi.ac.uk/pdbe/) ([PDBe-KB consortium, 2020](https://doi.org/10.1093/nar/gkz853); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.26ek1v)) and the Electron Microscopy Data Bank [EMDB](https://www.ebi.ac.uk/pdbe/emdb/) ([Lawson et al., 2011](https://doi.org/10.1093/nar/gkq880); in [FAIRsharing](https://dx.doi.org/10.25504/FAIRsharing.651n9j)), (ii) for the USA: [RCSB PDB](https://www.rcsb.org/) ([Berman et al., 2000](https://doi.org/10.1093/nar/28.1.235); in [FAIRsharing](https://dx.doi.org/10.25504/FAIRsharing.2t35ja)) and (iii) for Japan: [PDBj](https://pdbj.org/) ([Kinjo et al., 2017](https://doi.org/10.1093/nar/gkw962); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.rs2815)). 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](https://ghddi-ailab.github.io/Targeting2019-nCoV/) (in [FAIRsharing](https://fairsharing.org/biodbcore-001525)) and includes the following:
   1. Compound libraries including the [ReFRAME](https://reframedb.org/) compound library ([Janes et al., 2018](https://doi.org/10.1073/pnas.1810137115); in [FAIRsharing](https://fairsharing.org/biodbcore-001526)) (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. [Drug Discovery Cloud Computing System on Alibaba Cloud](https://www.alibabacloud.com/solutions/lifesciences-ehpc)
   3. Data mining and integration of historical drug discovery efforts against coronavirus (e.g., SARS/MERS) using artificial intelligence (AI) and big data
   4. Molecular chemical modelling and simulation data using computational tools.
   5. widely utilised in the nuclear community. The latest versions of the two nuclear reaction data libraries are [JEFF-3.3](https://www.oecd-nea.org/dbdata/jeff/) ([Cabellos et al., 2017](https://doi.org/10.1051/epjconf/201714606004); in [FAIRsharing](https://fairsharing.org/bsg-s001485)) and ENDF/B-VIII.0 ([Brown et al., 2018](https://doi.org/10.1016/j.nds.2018.02.001); in [FAIRsharing](https://fairsharing.org/bsg-s001484)) with a significant upgrade in data for a number of nuclides ([Carlson et al., 2018](https://doi.org/10.1016/j.nds.2018.02.002))
   6. Neutron scattering data are stored in the internationally-adopted [ENDF-6](https://www.oecd-nea.org/dbdata/data/manual-endf/endf102.pdf) format ([Brown et al., 2018](https://dx.doi.org/10.1016/j.nds.2018.02.001); in [FAIRsharing](https://fairsharing.org/bsg-s001484)) maintained by [CSEWG](https://www.nndc.bnl.gov/csewg/)
   7. Processed structural information is submitted in the [PDBx/mmCIF](http://mmcif.pdb.org/) format ([Fitzgerald et al., 2006](https://dx.doi.org/10.1107/97809553602060000745); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.fd28en)).
3. Molecular Dynamics (MD) simulations
   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](https://www.unidata.ucar.edu/software/netcdf/docs/netcdf_data_model.html); see also ([Goni et al., 2013](https://www.bsc.es/sites/default/files/public/life_science/molecular_modeling/d7.3_-_white_paper_on_standards_for_data_handling.pdf); in [FAIRsharing](https://fairsharing.org/bsg-s001210/))
   2. Refined structural models from experimental structural data using MD simulations are stored in [.pdb](https://www.rcsb.org/pdb/static.do?p=file_formats/index.jsp#pdb) format ([Bernstein et al., 1977](https://doi.org/10.1016/s0022-2836(77)80200-3); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.9y4cqw)).
4. Computer-aided drug design data
   1. Virtual screening results are stored in 3D chemical data formats, such as [.pdb](http://www.wwpdb.org/documentation/file-format) ([Bernstein et al., 1977](https://doi.org/10.1016/s0022-2836(77)80200-3); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.9y4cqw))
   2. Structural formulas either in SMILES ([Anderson et al., 1987](https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=2000CAUR.PDF); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.qv4b3c)) or [IUPAC International Chemical Identifier](https://iupac.org/who-we-are/divisions/division-details/inchi/) (InChI), and identified through InChIKey, a non-proprietary identifier for chemical substances ([Heller et al., 2015](https://doi.org/10.1186/s13321-015-0068-4); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.ddk9t9)).

**Proteomics**

For a curated list of relevant repositories for Proteomics see [FAIRsharing using the query ’proteomics’](https://fairsharing.org/search/?q=proteomics&content=biodbcore&name=&taxonomies=&organisations=&shortname=&description=&supportlinks=&licenses=&countries=&maintainers=&expanded_onto_domains=&expanded_onto_disciplines=&user_defined_tags=&record_id=&miriam_id=&search_state=hidden). The [ProteomeXchange Consortium](https://doi.org/10.25504/FAIRsharing.92dt9d) enables searches across the following deposition databases, following common standards

**Metabolomics**

For a curated list of relevant repositories in Metabolomics see [FAIRsharing using the query](https://fairsharing.org/search/?q=proteomics&content=biodbcore&name=&taxonomies=&organisations=&shortname=&description=&supportlinks=&licenses=&countries=&maintainers=&expanded_onto_domains=&expanded_onto_disciplines=&user_defined_tags=&record_id=&miriam_id=&search_state=hidden) [‘metabolomics’](https://fairsharing.org/search/?q=metabolomics&content=biodbcore&name=&taxonomies=&organisations=&shortname=&description=&supportlinks=&licenses=&countries=&maintainers=&expanded_onto_domains=&expanded_onto_disciplines=&user_defined_tags=&record_id=&miriam_id=&search_state=hidden).

**Lipidomics**

The recommended repository for Lipidomics data is [MetaboLights](https://www.ebi.ac.uk/metabolights) ([Haug et al., 2020](https://doi.org/10.1093/nar/gkz1019); in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.kkdpxe)).

**Social Science**

In Social Science, 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. COVID-19 related social science data may be shared in a generalist repository. If you use a general repository (e.g. [Figshare](http://figshare.com/) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.drtwnh)), [Dryad](http://datadryad.org/) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.wkggtx)), [Harvard Dataverse](https://dataverse.harvard.edu) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.t2e1ss)), [openICPSR](https://www.openicpsr.org/openicpsr/) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.nvwz0x)), [Zenodo](https://www.zenodo.org) (in [FAIRsharing](https://doi.org/10.25504/FAIRsharing.wy4egf)) 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](https://www.openicpsr.org/openicpsr/covid19)) accepts data from multiple domains (and formats) as a generalist repository, but because it is run by a social science data repository, ICPSR, it offers relevant domain repository benefits (e.g., curation by domain curators, restricted data dissemination options) and ensures social science COVID-19 data are FAIR and in a CoreTrustSeal repository.

### Publications / Data Publications

Rapid publications, 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 available immediately upon communication of results, e.g. through a press release.

Research funders and policy makers should implement a “data first” publication policy by encouraging the publication of data articles in “open” peer-reviewed data journals, or mandating and supporting the deposit of data and associated code in a trustworthy data repository in tandem with the publication of articles. Curated datasets and peer-reviewed data 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 data underlying a study, in an even more timely manner than usual. Publishers need to make sure that the author recommendations prefer publishing of data in trustworthy domain-specific repositories where findability is better than in generic or institutional repositories.

### D. Software

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.

*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)](https://doi.org/10.7717/peerj-cs.86).

Examples: The FORCE11 Software Citation Implementation Working Group [(Chue Hong et al., 2017)](https://www.force11.org/group/software-citation-implementation-working-group) 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](https://journals.aas.org/software-citation-suggestions/)

*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)](https://zenodo.org/) and Figshare [(FigShare, 2020)](https://figshare.com/). 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)](https://doi.org/10.17605/OSF.IO/KDE56). If the software is deposited along with data ([DataCite, 2020](https://doi.org/10.17616/R3D)), 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)](https://joss.theoj.org) 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)](https://www.force11.org/group/software-citation-implementation-working-group). The GigaScience journal is another example of publication requiring the availability of software ([GigaScience, 2020](https://academic.oup.com/gigascience/pages/instructions_to_authors)).

*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)](https://joss.theoj.org); SoftwareX submission process includes two mandatory metadata tables that include licence and code availability [(Elsevier, 2020)](https://www.journals.elsevier.com/softwarex).

1. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)