



OPEN LETTER

Fostering global data sharing: highlighting the recommendations of the Research Data Alliance COVID-19 working group [version 1; peer review: 1 approved, 2 approved with reservations]

Claire C. Austin ¹, Alexander Bernier ², Louise Bezuidenhout ³, Juan Bicarregui ⁴, Timea Biro ⁵, Anne Cambon-Thomsen ⁶, Stephanie Russo Carroll ⁷, Zoe Cournia ⁸, Piotr Wojciech Dabrowski ⁹, Gayo Diallo¹⁰, Thomas Duflot ¹¹, Leyla Garcia ¹², Sandra Gesing¹³, Alejandra Gonzalez-Beltran ⁴, Anupama Gururaj ¹⁴, Natalie Harrower ⁵, Dawei Lin¹⁴, Claudia Medeiros ¹⁵, Eva Méndez ¹⁶, Natalie Meyers ¹⁷, Daniel Mietchen ¹⁸, Rajini Nagrani ¹⁹, Gustav Nilsson ²⁰, Simon Parker²¹, Brian Pickering ²², Amy Pienta²³, Panayiota Polydoratou ²⁴, Fotis Psomopoulos ²⁵, Stephanie Rennes²⁶, Robyn Rowe ²⁷, Susanna-Assunta Sansone ²⁸, Hugh Shanahan ²⁹, Lina Sitz ³⁰, Joanne Stocks ³¹, Marcos Roberto Tovani-Palone^{32,33}, Mary Uhlmansiek ³⁴,
Research Data Alliance

¹Environment and Climate Change Canada, 351 boul. St-Joseph, Gatineau, Quebec, K1A 0H3, Canada

²Centre of Genomics and Policy, McGill University, 740, avenue Dr. Penfield, suite 5200, Montreal, Quebec, Canada

³Institute for Science, Innovation and Society, University of Oxford, 64 Banbury Road, Oxford, OX2 6PN, UK

⁴UKRI-STFC Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK

⁵Digital Repository of Ireland, Royal Irish Academy, 19 Dawson St, Dublin 2, D02 HH58, Ireland

⁶CNRS, Inserm and University Toulouse III Paul Sabatier, Toulouse, France

⁷Native Nations Institute at the Udall Center for Studies in Public Policy and the College of Public Health, University of Arizona, 803 E First ST, Tucson, AZ, 85719, USA

⁸Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, Athens, 11527, Greece

⁹HTW Berlin University of Applied Science, Wilhelmshofstraße 75A, Berlin, 12459, Germany

¹⁰BPH INSERM1219 & LaBRI, Univ. Bordeaux, 146 rue Léo Saignat, F-33000, Bordeaux, France

¹¹Normandie Univ, UNIROUEN, CHU Rouen, Department of Clinical Research, Rouen University Hospital, 1 Rue de Germont, Rouen Cedex, 76031, France

¹²ZB MED Information Centre for Life Sciences, Gleueler Str 60, Cologne, 50931, Germany

¹³University of Notre Dame Center for Research Computing, 814 Flanner Hall, Notre Dame, IN, 46556, USA

¹⁴National Institute of Allergy and Infectious Diseases, National Institutes of Health, 5601 Fishers Lane, Rockville, MD, 20852, USA

¹⁵Institute of Computing, University of Campinas, Av Albert Einstein 1251, Campinas, São Paulo, 13082-853, Brazil

¹⁶Universidad Carlos III de Madrid, C/ Madrid, 128, Getafe (Madrid), 28903, Spain

¹⁷250D Navari Center for Digital Scholarship, Hesburgh Library, University of Notre Dame, Notre Dame, IN, 46556, USA

¹⁸School of Data Science, University of Virginia, P.O. Box 400249, Charlottesville, VA, 22904, USA

¹⁹Leibniz Institute for Prevention Research and Epidemiology, Achterstrasse 30, Bremen, 28359, Germany

²⁰Karolinska Institutet & Swedish National Data Service, Nobels väg 9, Stockholm, 17177, Sweden

²¹Cancer Research UK, 2 Redman Place, London, E20 1JQ, UK

²²University of Southampton, University Road, Southampton, SO17 1BJ, UK

²³ICPSR, University of Michigan, P.O. Box 1248, Ann Arbor, MI, 48106-1248, USA

²⁴OpenEdition/Department of Library Science, Archives and Information Systems, International Hellenic University, P.O. Box 141, Thessaloniki, 57400, Greece

²⁵Institute of Applied Biosciences (INAB), Centre for Research and Technology Hellas (CERTH), Thessaloniki, 57001, Greece

²⁶INRAE National Research Institute for Agriculture, Food and Environment, 147 Rue de l'Université, Paris, 75007, France

²⁷Laurentian University, Ontario, P3E 2C6, Canada

²⁸Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, Oxford, OX1 3QG, UK

²⁹Department of Computer Science, Royal Holloway, University of London, Bedford Building, Egham, TW20 0EX, UK

³⁰Independent Researcher, Strada Costiera, Trieste, 34151, Italy

³¹Division of Rheumatology, Orthopedics and Dermatology, School of Medicine, University of Nottingham, Queens Medical Centre, Nottingham, NG7 2UH, UK

³²Ribeirão Preto Medical School, University of São Paulo, São Paulo, Brazil

³³Modestum Ltd, Hilton, Derbyshire, UK

³⁴Research Data Alliance - US Region (RDA-US), c/o Ronin Institute, 127 Haddon Place, Montclair, NJ, 07043, USA

V1 First published: 09 Nov 2020, 5:267
<https://doi.org/10.12688/wellcomeopenres.16378.1>

Latest published: 09 Nov 2020, 5:267
<https://doi.org/10.12688/wellcomeopenres.16378.1>

Abstract




The systemic challenges of the COVID-19 pandemic require cross-disciplinary collaboration in a global and timely fashion. Such collaboration needs open research practices and the sharing of research outputs, such as data and code, thereby facilitating research and research reproducibility and timely collaboration beyond borders. The Research Data Alliance COVID-19 Working Group recently published a set of recommendations and guidelines on data sharing and related best practices for COVID-19 research. These guidelines include recommendations for researchers, policymakers, funders, publishers and infrastructure providers from the perspective of different domains (Clinical Medicine, Omics, Epidemiology, Social Sciences, Community Participation, Indigenous Peoples, Research Software, Legal and Ethical Considerations). Several overarching themes have emerged from this document such as the need to balance the creation of data adherent to FAIR principles (findable, accessible, interoperable and reusable), with the need for quick data release; the use of trustworthy research data repositories; the use of well-annotated data with meaningful metadata; and practices of documenting methods and software. The resulting document marks an unprecedented cross-disciplinary, cross-sectoral, and cross-jurisdictional effort authored by over 160 experts from around the globe. This letter summarises key points of the Recommendations and Guidelines, highlights the relevant findings, shines a spotlight on the process, and suggests how these developments can be leveraged by the wider scientific community.


Keywords

Open science, Sharing research outputs in pandemics caused by infectious diseases, FAIR and CARE principles, Omics, Epidemiology, Social Science, Clinical Research, COVID-19

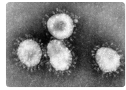
Open Peer Review

Reviewer Status 

	Invited Reviewers		
	1	2	3
version 1			
09 Nov 2020	report	report	report

- Elena Simperl**, King's College London, London, UK
- Fruzsina Molnar-Gabor**, Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany
- Maria Luiza Machado Campos** , Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the [Coronavirus \(COVID-19\)](#) collection.

Corresponding author: Research Data Alliance (secretariat@rda-foundation.org)

Author roles: **Austin CC:** Conceptualization, Resources; **Bernier A:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Bezuidenhout L:** Writing – Original Draft Preparation, Writing – Review & Editing; **Bicarregui J:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Biro T:** Conceptualization, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; **Cambon-Thomsen A:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Carroll SR:** Writing – Original Draft Preparation; **Cournia Z:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Dabrowski PW:** Conceptualization, Writing – Original Draft Preparation; **Diallo G:** Conceptualization, Writing – Review & Editing; **Duflot T:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Garcia L:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Gesing S:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Gonzalez-Beltran A:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Gururaj A:** Conceptualization, Writing – Review & Editing; **Harrower N:** Writing – Original Draft Preparation, Writing – Review & Editing; **Lin D:** Conceptualization, Writing – Review & Editing; **Medeiros C:** Writing – Original Draft Preparation; **Méndez E:** Conceptualization; **Meyers N:** Conceptualization, Writing – Original Draft Preparation; **Mietchen D:** Writing – Review & Editing; **Nagrani R:** Writing – Original Draft Preparation; **Nilsonne G:** Writing – Review & Editing; **Parker S:** Conceptualization; **Pickering B:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Pienta A:** Writing – Original Draft Preparation, Writing – Review & Editing; **Polydoratou P:** Writing – Original Draft Preparation; **Psomopoulos F:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Rennes S:** Writing – Original Draft Preparation; **Rowe R:** Conceptualization, Writing – Review & Editing; **Sansone SA:** Conceptualization, Resources, Writing – Review & Editing; **Shanahan H:** Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Sitz L:** Writing – Original Draft Preparation, Writing – Review & Editing; **Stocks J:** Writing – Original Draft Preparation, Writing – Review & Editing; **Tovani-Palome MR:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Uhlmansiek M:** Project Administration, Resources, Writing – Review & Editing;

Competing interests: No competing interests were disclosed.

Grant information: The Research Data Alliance has been supported through a grant from the Open Research division of Wellcome. A. Barnier is supported by the Canadian Cancer Genome Collaboratory. A. Cambon-Thomsen and E. Méndez are supported by RDA EU 4.0 project, ref. GA no. 777388. Z. Cournia is supported by the European Unions Horizon 2020 European research infrastructures, "National Initiatives for Open Science in Europe – NI4OS Europe" project (grant agreement no. 857645). G. Diallo is supported by AUF Europe de l'Ouest (Patient-Covid19 project) and Univ. Bordeaux (France). C. Medieros is supported by FAPESP - Sao Paulo State Foundation Grant 13/08293-7. A. Pienta is supported through the Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan and National Institute on Drug Abuse 75N95019C00017. S. Sansone is supported through Wellcome grant 212930. M. Uhlmansiek is supported through US National Science Foundation grant 1934649.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Austin CC *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

How to cite this article: Austin CC, Bernier A, Bezuidenhout L *et al.* **Fostering global data sharing: highlighting the recommendations of the Research Data Alliance COVID-19 working group [version 1; peer review: 1 approved, 2 approved with reservations]** Wellcome Open Research 2020, 5:267 <https://doi.org/10.12688/wellcomeopenres.16378.1>

First published: 09 Nov 2020, 5:267 <https://doi.org/10.12688/wellcomeopenres.16378.1>

Disclaimer

The views expressed in this article are those of the author(s). Publication in Wellcome Open Research does not imply endorsement by Wellcome.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic is currently one of the most challenging global issues, with economic, social, political, cultural and scientific consequences (Nicola *et al.*, 2020; Rajkumar, 2020). The rapid spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus and the need for global stewardship has led researchers to collaborate on a worldwide scale, escalating the production of scientific data and highlighting the urgency to provide those data in an accessible, re-usable, and timely manner.

To ensure the rapid sharing of high-quality data, the [Research Data Alliance \(RDA\)](#) established a [rapid-response working group on COVID-19](#), which quickly grew to more than 600 members, with over 160 individuals contributing actively to recommendations over a 10 week period. The working group was 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).

The objective of this RDA working group, in the first instance, was to provide data sharing recommendations for researchers, clinicians, policymakers, funders, publishers, and providers of data sharing infrastructures concerning the most important challenges raised by the current pandemic.

The final version of the RDA COVID-19 Recommendations and Guidelines on Data Sharing ([RDA COVID-19 WG, 2020](#)) was released on 30th June 2020 and provides up-to-date advice across the eight areas mentioned above to support robust and meaningful data reuse for the COVID-19 pandemic management. Each sub-section of the 143-page document is organised into four main subparts: “*Focus and Description*”, “*Scope*”, “*Policy recommendations*” and “*Guidelines*”, allowing efficient navigation for the reader seeking precise information.

Prefaced by an executive summary, it provides essential reference text for relevant stakeholders, with granular guidelines for researchers and data managers preceded in each section by higher level recommendations to policymakers and funders. It is also connected to an extensive bibliography accessible via the participative online service Zotero Library ([RDA COVID-19 Zotero WG, 2020](#)). An [infographic](#) was created to provide an overview and highlight key areas. Other resources, in particular a decision-making tool and a mindmap are in development so that readers can more efficiently navigate the document. These are being made available on the [Value of RDA for COVID-19](#) webpage. The comprehensive recommendations and related navigation tools facilitate uptake by all stakeholders (including the public), who wish to access and contribute reliable information on the global COVID-19 research and response process.

Since disciplines and communities often develop *ad hoc* data management practices that are prone to becoming siloed, the report encourages data exchange between stakeholders. It highlights the advances and procedures in different disciplines, but crucially also draws attention to the commonalities between disciplines, fostering interdisciplinary action, understanding of the disciplines that stakeholders are not part of, and future collaboration. The RDA is in a unique position to develop such guidance due to its grassroots, participative tradition of interdisciplinary self-motivated dialogue and solutions-based outputs.

Recommendations

In this section, we provide a brief motivation for each subgroup, the problems identified, and a summary of key recommendations, per group as well as overarching guidance.

Clinical guidelines

Healthcare measures and clinical research are at the forefront of combating the COVID-19 pandemic. Obtaining actionable clinical information about the disease and seeking an effective treatment to fight the infection are key to minimising the impact of this unprecedented global health challenge. Clinical trials should follow the International Council for Harmonisation (ICH) efficacy guidelines to ensure the data quality. As cases rise, the promotion of clinical data sharing is of utmost importance. Many studies and trials are performed under enormous time pressure, which can weaken the methodology and lead to preliminary results being published without a full review. We recommend making the data behind research available alongside research results. The recommendations detail how to use trustworthy repositories to provide transparency, integrity and context to data for timely discovery and the validation of new findings. A key goal is to avoid policy-making based on fraudulent studies, which in turn causes distrust in science ([Group, The Editors of the Lancet, 2020](#)).

Omics guidelines

Omics-scale studies of SARS-CoV-2 are emerging rapidly with exceptional potential to unravel the mechanisms of the COVID-19 pathobiology. These studies offer new mechanistic insights into the pathogenesis of COVID-19 and ways forward for diagnostic and therapeutic intervention, while at the same time generating a tremendous amount of data. The Omics subgroup was motivated to draft guidelines based on the requirement for rapid, open data sharing. This rapid sharing facilitates early insights into the molecular biology of the COVID-19 processes at a cellular level, possibly leading to new therapeutic targets, diagnostic markers and disease management. Omics research should be a collaborative effort to learn the genetic determinants of COVID-19 susceptibility, severity and outcomes. Thus, the use of domain-specific repositories to enable standardisation of terms and enforce metadata standards is mandated. Availability and re-usability of research data on COVID-19 in order to prevent unnecessary duplication of work is described for virus genomics, host genomics, proteomics, metabolomics, lipidomics, and structural data. The efforts of the RDA Omics working group provides clear recommendations of repositories to find

existing data depending on the target methodology in the above research areas, as well as best practices for sharing data and identifying the most prevalent data and metadata formats.

Epidemiology guidelines

An immediate understanding of the COVID-19 epidemiology is crucial to slowing infections, minimising deaths, making informed decisions about when, and to what extent, to impose mitigation measures, and when and how to reopen society. One of the major challenges encountered in the field of epidemiology is that the data and models are uncomparable, frequently incomplete, provisional, and subject to correction under changing conditions, making their use and reuse for timely epidemiological analysis challenging. The principal guidelines for researchers are to ensure that the data models must be inclusive of not only clinical data, disease milestones, indicators and reporting data, but also contact tracing and personal risk factors. Our recommendations for the policymakers are to incentivise the publication of situational data, analytical models, scientific findings, and reports used in decision making.

Social sciences guidelines

The social sciences recommendations seek to ensure that social science data is widely (re)usable to answer fundamental questions about social aspects of the pandemic and that the data are accessible for work ongoing in other domains. The subgroup recommendations include: encouraging data management that follows best practices and improves data sharing; use of trustworthy repositories to share data; retention of information (e.g., geographic information) to allow data linkage within and across domains while maintaining confidentiality; access to measures that are useful when making statistical adjustments for selection bias, thereby improving the representativeness of findings from limited samples; and balancing the desire to share data widely with ensuring that human subjects protections are met and confidential data are kept secure.

Community participation guidelines

Community participation guidelines were created with the aim of bridging stakeholder involvement; ensuring that inputs from researchers, citizen scientists, developers and device makers are streamlined, with perspectives from patients, policymakers and the public at large also considered. Linking communities and supporting communication will help coordination and avoid duplication of efforts since many communities are driving similar or complementary efforts in response to the current public health emergency. These recommendations aim to support the varied work of communities in sharing data to improve research outputs and public knowledge.

Guidelines for data sharing respecting indigenous data sovereignty

Indigenous Peoples and nations globally need to be actively engaged in governance processes that include Indigenous-related COVID-19 data, data lifecycles, and data ecosystems. This is a necessary part of respecting the inherent rights of Indigenous nations to have sovereignty and governance over Indigenous data. The Indigenous COVID-19 data guidelines set out the minimum requirements for Indigenous-designed data approaches

for funders, governments, researchers, and data stewards regarding governance, collection, ownership, application, sharing, and dissemination of Indigenous data, specifically in relation to COVID-19. These guidelines reflect and support Indigenous Data Sovereignty (see www.GIDA-global.org), underpinned by the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP) and framed around the CARE (for Collective Benefit, Authority to Control, Responsibility, Ethics) Principles for Indigenous Data Governance (RDA, 2018). These guidelines do not supersede or replace existing Indigenous governance protocols or agreements developed (or under development) by Indigenous Peoples or nations. Rather, they 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, so they are aligned to ethical and cultural Indigenous data practices supported by collective consent. This demands proactive investment in Indigenous community-controlled data infrastructures to support community capacity and resilience, and improve the flow of information for effective public health response.

Software guidelines

Regardless of the research domain, software plays a fundamental role to realise reproducible science as it enables analyses and processing of data. The recommendations for research software covers aspects of development, release and maintenance, derived from previous work (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). Our recommendations to researchers focus on key practices enabling (re)use of research software making it easier for other researchers to build upon and focus their efforts on new approaches. Openness, availability, documentation and examples are key elements here. Before software is re-used, it must be found; therefore, our recommendations focus on software citation, archives and deposit platforms for released versions and alignment with publishing best practices. Finally, neither software development nor its publication are possible without sufficient funding support. In this sense, we centred our recommendations on increasing the recognition of software, its role in reproducibility, and funding opportunities not only for development but also for maintenance and sustainability.

Legal and ethics guidelines

Data sharing must occur in compliance with relevant legal and ethical frameworks. The legal and ethics section describes sources of legal and ethical obligations applicable to researchers performing studies on COVID-19, including biomedical and social science research ethics guidance. The section makes recommendations to help ensure best practices are respected in using COVID-19 data across jurisdictions and institutions. Recommendations include a synthesis of foundational principles of data privacy in law and ethics, and a description of organisational data governance practices. The recommendations also help researchers identify the most appropriate actor at their institution to guide them in adhering to local legal and ethical requirements. Data governance is considered throughout the data lifecycle in the spirit of community engagement and benefit sharing. Best practices for data de-identification and anonymisation, as well as data and software IP licensing are described. A discussion

of the distinct consent standards applicable to clinical care, research ethics, and data privacy law is incorporated.

Overarching recommendations/foundational elements

In addition to each group's recommendations, the document starts with a series of overarching recommendations. These foundational elements draw directly from the findings of the subgroups, as well as from broader current discussions on research data sharing and Open Science, tailored to the critical need for timely, precise, and interoperable research data sharing under a pandemic.

The sharing of research data promotes research integrity, enables others to investigate results, and fosters the very purpose of research itself - to build upon existing knowledge towards new discoveries. The timely sharing of well-curated data (and software, algorithms, and other resources) enables reuse, often for purposes unanticipated by the research that first produced the data. For this reuse to be possible, data must be collected, documented, curated, preserved, and made available through trusted and recognised platforms. The FAIR principles (Wilkinson *et al.*, 2016) - promoting data to be Findable, Accessible, Interoperable and Reusable - provide a well-recognised framework for data sharing, and were noted frequently by contributors across the sections.

Disciplinary borders provide one challenge, but so do geographical and administrative boundaries. COVID-19 does not respect borders of any kind, so, similarly, neither can research. The need for cross-jurisdictional efforts to support sharing of data and other resources, through coordination, funding and legal agreements, is also key. Computational infrastructures need to be refreshed and invested in as a public good; investment in technology needs to be accompanied by support for human resources to maintain infrastructure, and training programmes in data stewardship need to be developed and offered broadly. Data and other outputs need to be prepared for sharing so that they are understandable, and this process should be started as early as possible in the research process with the creation of a data management plan (DMP), which details how data will be stewarded throughout the research lifecycle. This lifecycle is key to the remaining 'Foundational' elements: data must be accompanied by documentation such as research methods, context, data manipulation; rich metadata in standard formats need to accompany outputs, data should be deposited in domain-suitable trustworthy data repositories for discovery, preservation and reuse, and the rapid publication of data should be encouraged supported, and mandated by funders and publishers.

Discussion

A key aim of the recommendations and guidelines has been to offer both system-wide and concrete guidance to facilitate data sharing across disciplines and geographical boundaries in a timely and accurate manner, thus helping accelerate the time to a cure, supporting informed decisions and improving the global response to the pandemic.

The involvement of specialists and practitioners coming from the many disciplines and fields impacted by the pandemic has ensured that the report is both expert-informed, as well as

community reviewed. The incorporation of repeated open consultations was also meant to facilitate a fast track path to wider adoption, considering that researchers, policymakers and other stakeholders have been involved as early as possible in the formulation, drafting, and consensus on the document. The priority is for these guidelines and recommendations to be widely adopted in order to accelerate solutions to the pandemic.

Instead of a silo-based approach, the document points out the commonalities in data management across different research areas and themes. Identifying commonalities implies that similar solutions can be identified and applied. This bridge from the STEM (Science, Technology, Engineering, and Maths) to social science aspects of the COVID-19 challenge demonstrated how truly interdisciplinary work across fields can provide valuable insights and stimulate a creative process. The added value of such overarching cooperations is a key takeaway from this process that is also likely to enrich similar efforts.

The document was developed with a comparatively light level of moderation and emerged on a very rapid time-frame of 10 weeks, including the release of five drafts for open consultation. Writing coordination focused on ensuring the flow of information, so the sub-groups met regularly as did the moderators and chairs. There was a weekly webinar open to the public, as well weekly Co-Chairs meetings, and weekly coordination sessions for Chairs and Moderators; in addition, each subgroup agreed their own meeting plans. Small teams were set up for visualisation of recommendations, and for managing references. The foundational elements and executive summary were drafted by the editorial team, undergoing successive editing phases, where participants from different groups could comment widely across the whole document. This light-weight structure was enabled through relatively simple tools, namely Google Docs, Zotero and conference calls. The final publication is designed as a reference text, where readers are likely to selectively read parts of the document relevant to them, so a certain degree of repetition on key advice was retained to address this selective reading.

Going forward, the RDA COVID-19 initiative has shown that there is a global willingness among experts from a range of disciplines to engage with the grand challenges we face as well as to generously offer their time and experience to generate thorough and well-rounded guidance that is attentive to philosophical and pragmatic differences. This experience made clear that to a great extent, the knowledge, expertise, and solutions for working together in the face of global emergencies is already in place, so we need to foster this through continued coordination, harmonisation, and decision making.

Conclusions

The report has highlighted the importance of data sharing in different domains with respect to COVID-19. It provides a range of detailed guidelines aimed at communities with different practices of data management. The guidelines directly target researchers to facilitate best practices and maximise efficiency while also addressing policy makers, funders, publishers and providers of data infrastructures with a framework for future emergencies. With over 600 members, the group reached a

substantial size with diverse knowledge, background and domain experience.

As reflected in the content of this paper, going forward, the RDA COVID-19 WG is not only focused on the wider communication and adoption of the recommendations and guidelines themselves but also on providing best practices for the process of developing similar reports and outputs in the context of a multi-disciplinary, bottom-up and geographically diverse community, to be able to answer global rapid challenges like the COVID-19 pandemic.

The RDA is engaging with stakeholders at various levels to build impact and encourage adoption of the guidelines. From a policy perspective, the WG was instigated rapidly in response to a request by the European Commission and the guidelines are an important resource for the Organisation for Economic Co-operation and Development (OECD), the Bill and Melinda Gates Foundation, and the Wellcome Trust. From a research perspective, the work is continuing through a number of discipline-specific research papers, domain-focused narratives, and use-cases, for example, within the epidemiology and legal/ethical subgroups. Furthermore, the process of forming the collaboration, developing the guidelines and quantifying their impact, is being analysed as a piece of social science research in its own right. Other group members are pursuing extended efforts through the creation of new interest or working groups through the RDA that will meet for the first time at virtual plenary in November 2020. The RDA itself is committed to sharing and improving the approach as an example of good practice, offering its structure and overall processes support as a framework for similar efforts.

The experience of writing the guidelines shows that the creation of a document with contributions from a large, diverse group is possible in a relatively short amount of time. Subgroups can operate in tandem to save time; however, they require editors to

move different sections towards completion, and to help create a consistent structure and approach throughout the final document. A framework to steer the subgroups towards a common goal, particularly in terms of the intended audience, is also crucial. This community-driven writing can serve as a template for future world-wide urgent challenges such as the next pandemic, a natural disaster or indeed the climate crisis. The urgency and unprecedented global and near simultaneous nature of the pandemic likely contributed to participant motivation, and the question remains of how similarly large scale, multidisciplinary challenges may be addressed when the urgency is not as palpable. Without such urgency, this might attract fewer contributors. Nevertheless, as described here, this still provides a good mechanism for creating key guidelines that reflects a large diverse community.

Data availability

No data are associated with this article.

Author contributions

The roles of each of the contributing authors is described using the CRediT Taxonomy [here](#). The list of authors above is strictly alphabetical. The final author, the RDA indicates the contribution of the organisation to this paper.

Acknowledgements

This work was developed as part of the 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. We acknowledge specifically the contributors to the document 'RDA COVID-19 Recommendations and Guidelines on Data Sharing' and the members of the RDA-COVID19 WG.

References

- Akhmerov A, Cruz M, Drost N, *et al.*: **Raising the Profile of Research Software**. *Zenodo*. 2019.
[Publisher Full Text](#)
- Anzt H, Bach F, Druskat S, *et al.*: **An Environment for Sustainable Research Software in Germany and beyond: Current State, Open Challenges, and Call for Action [version 1; peer review: 1 approved, 1 approved with reservations]**. *F1000Res*. 2020; 9: 295.
[Publisher Full Text](#)
- Clément-Fontaine M, Di Cosmo R, Guerry B, *et al.*: **Encouraging a Wider Usage of Software Derived from Research**. Research Report. Committee for Open Science's Free Software and Open Source Project Group, 2019.
[Reference Source](#)
- Jiménez RC, Kuzak M, Alhamdoosh M, *et al.*: **Four Simple Recommendations to Encourage Best Practices in Research Software [version 1; peer review: 3 approved]**. *F1000Res*. 2017; 6: ELIXIR-876.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lamprecht AL, Garcia L, Kuzak M, *et al.*: **Towards FAIR Principles for Research Software**. *Data Sci. Preprint*, 2019; 3(1): 37–59.
[Publisher Full Text](#)
- Nicola M, Alsafi Z, Sohrabi C, *et al.*: **The socio-economic implications of the coronavirus pandemic (COVID-19): A review**. *Int J Surg*. 2020; 78: 185–193.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rajkumar RP: **COVID-19 and mental health: A review of the existing literature**. *Asian J Psychiatr*. 2020; 52: 102066.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- RDA: **CARE Principles of Indigenous Data Governance**. 2018.
[Reference Source](#)
- RDA COVID-19 Working Group: **RDA COVID-19 Recommendations and Guidelines on Data Sharing**. Research Data Alliance. 2020.
[Publisher Full Text](#)
- RDA-COVID19 Zotero WG: **RDA-COVID19 Zotero Library**. Research Data Alliance. 2020.
[Publisher Full Text](#)
- The Editors of the Lancet Group: **Learning from a retraction**. *Lancet*. 2020; 396(10257): 1056.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **The FAIR Guiding Principles for Scientific Data Management and Stewardship**. *Sci Data*. 2016; 3: 160018.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wilson G, Bryan J, Cranston K, *et al.*: **Good Enough Practices in Scientific Computing**. *PLoS Comput Biol*. 2017; 13(6): e1005510.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status: ? ? ✓

Version 1

Reviewer Report 13 January 2021

<https://doi.org/10.21956/wellcomeopenres.18015.r41351>

© 2021 Campos M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Maria Luiza Machado Campos 

Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

The article presents an overview of the results from the COVID-19 Research Data Alliance Working Group, originally delivered in a report addressing the challenges of data sharing during the COVID-19 pandemic. In that report, the working group discussed a series of recommendations and guidelines divided into four research areas and four cross-cutting themes. The article highlights the key points and main findings associated to these guidelines, also describing their development process.

As the original WG report is a long and comprehensive document, the article is an important and useful reading. It is not meant to replace the original report, however it gives a general view of the whole content, serving to increase its dissemination in the scientific community and to give a more tangible understanding to those who do not belong to the specific addressed domains or even for those who need to engage in similar initiatives.

The introduction is very motivating, presenting the working group and the process of designing guidelines and recommendations. It gives an idea of the effort made during development, especially if we consider the number of participants and the short time on which the report was produced. It also includes links to an overview infographic and to other nice associated resources complementary to the report and to the article.

At the recommendations section, the authors refer to each of the domain and crosscutting themes with their specific requirements and assumptions and complement them with common issues. Perhaps this section could benefit from better balancing the content of the different subgroups or areas. Some are very well described, with a good presentation of results (such as Software guidelines, Omics guidelines), while others, in contrast, are simpler (such as Clinical Guidelines and Community participation guidelines). Community participation, for example, could have its relevance and characterization better justified and described, so that the readers could better understand its priority over other candidate themes.

During the discussion session, it would be interesting to have some comparison or comments in

relation to other existing works addressing directions for the COVID-19 research projects or more general data strategies, as many of these topics have been highly debated in the academic and government areas. Maybe the joint effort Data Together, involving RDA, CODATA, GO FAIR and EOSC could have been mentioned. Although the FAIR principles have been cited, initiatives like GO FAIR VODAN IN (the Virus Outbreak Data Network) were not discussed. Such works could certainly complement the RDA WG results and serve as further references for the readers.

Obs: The article is well presented and written, but please correct "The recommendations for research software COVER aspects..." (and not COVERS).

Is the rationale for the Open Letter provided in sufficient detail?

Yes

Does the article adequately reference differing views and opinions?

Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Yes

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Data Management, Information integration, Ontology engineering.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 January 2021

<https://doi.org/10.21956/wellcomeopenres.18015.r41353>

© 2021 Molnar-Gabor F. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Fruzsina Molnar-Gabor

Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany

The article is a timely roadmap for communally and consistently dealing with the crucial global

challenges arising from the COVID-19 pandemic (as they pertain to data sharing). It delivers substantive cross- and inter-disciplinary guidance and recommendations for those stakeholders able to influence the international response to these challenges. That the RDA was able to deliver such a comprehensive report within the extremely limited timeframe of ten weeks is down to just how well-situated the Alliance is to take up the function of organizing and establishing interdisciplinary responses to global challenges and how well-placed it is – through its organizational and management efforts and its access to and use of relevant expertise – to deliver substantial answers to thorny questions arising from the current crises. On the other hand, even the convenient situation and placement of the RDA would not have been enough to respond adequately without its determined work early in the pandemic to establish a working group on COVID-19, which was crucial to shaping the RDA's response.

Indeed it is this very procedure – of developing an adequate response in the form of recommendations and guidelines to foster data sharing – that is most relevant for both the current pandemic and any potential future global challenges of comparable nature, as these require solutions which transcend individual communities and disciplines and can only be reached by joining forces. While respecting and appreciating the knowledge and professional expertise of all members that have contributed to this remarkable result, it is only the final presentation of the results of the report in the article and not the substantive content of the report itself that can be subject to this present review.

Thus, this review is motivated by the assumption that readers will most probably read the article before engaging with the whole report. Accordingly, this review aims to ensure the article provides a clear representation of the results. Additionally, while the report has an Executive Summary which should not be replaced by the article, the article would be a good place to elaborate on the implementation and application of the guidance found within the report.

To begin the review, I would first like to address the structure and main terms of the article as they relate to the structure of the report:

Firstly, the article presents a good summary of the results of the working group in a manner likely to engage readers and encourage them to read the main report, as it points out structural-substantive elements that could draw readers' interest.

While the presentation of the structural parts is well-balanced, a more detailed presentation of the four cross-cutting themes would be appreciated, as these are of interest for all four research areas. A more detailed summary would benefit readers of all disciplines, interests and purposes.

Secondly, there are minor inconsistencies in the structure of the document that should be remedied. For example, the title of the 'Recommendations' section is slightly confusing given the section relates to both recommendations for policymakers and funders and guidelines for researchers and clinicians. Additionally, the name of the 'Overarching Recommendations' subsection is confusing due to the overlap of the terms used, not just in terms of vocabulary but also metatextually in how it alters the presentation of the results. Adding the specific level of recommendations (such as foundational-overarching, area-specific or both) to the subtitles in the relevant subsections would help readers quickly assess the structure.

Thirdly, there is no substructure to help differentiate between the guidance in each of the four research areas and those relating to the overarching themes, so the 'Recommendations' section

could also be adapted to mirror the main structure of the report here.

The second set of suggestions relates to the usage of certain terms that play a substantive role in the article.

Firstly, the 'Introduction' section and the abstract name different addressees (clinicians are missing in the abstract). It is always valuable to include the full list of addressees wherever possible, particularly in the abstract, not just to provide readers with complete information but also to attract relevant stakeholders.

Secondly, "data sharing", "data reuse" and "data exchange" seem to be used in the text as equivalent terms but in many jurisdictions – and when examined ethically – they are not entirely interchangeable. This is noted less to spark an academic discussion about interpretative approaches to these notions or about partially-missing binding legal definitions, but more due to the fact that this vocabulary is central to the content as a whole and should, for the purposes of this article, be further harmonized.

Thirdly, "data managers" seems to be being used as an overarching term for selected addressees. Whether this is in fact the case and how the term and the group of actors included relate to the other, explicitly highlighted groups of addressees needs to be clarified.

Fourthly, technical and legal/normative interoperability should be – where relevant – labelled as such.

The third set of suggestions relates to the presentation of the guidance.

Firstly, in the area of social sciences, the guidelines and recommendations are more strongly connected than in other areas due to the characteristics of this field. This could be emphasized in the article, as is also recommended by the detailed presentation in the report.

Secondly, in the section 'Legal and ethics guidelines', as well as elsewhere in the article, legal challenges could be better emphasized in general to motivate legislators and relevant policymakers to push for frameworks for open infrastructures and rules for the legal securing of data sharing. This is especially relevant in a situation where weighing contradictory legal and ethical positions – usually so difficult in the traditional setting of health data sharing – is already gradually becoming easier and where the pandemic is now forcing a remarkable shift towards an overlap between originally contradictory poles. Emphasizing addressees' interests is also crucial for grounding their legal positions and the weighing of interests will always ultimately also be guided by those of the public and society (similarly indicated in the report itself, cf. Executive Summary, subsection 'Recommendations', p. 9; Section 6. Data Sharing in Social Sciences, subsection 6.4.4, p. 47; Section 10. Legal and Ethical Considerations, subsection 10.2, p. 67).

Thirdly, the section on legal and ethical guidelines could be more clearly structured to summarize and separate recommendations for policy makers and other related addressees (such as providing the conditions for the relevant actors to be able to work according to the FAIR principles), and guidelines for researchers and other related addressees. This seems particularly important as the report itself emphasizes the role of law (as related to open science through policy: Executive Summary, subsection 'Recommendations', p. 8; related to the implementation of legal frameworks

that promote sharing of data across jurisdictions and sectors: 10. Legal and Ethical Considerations, subsection 10.3.1, Nr. 9, p. 9).

The fourth set of suggestions concerns the 'Discussion' and 'Conclusions' sections.

The authors consistently and correctly use "best practices" in the plural, clearly recognizing that the developed recommendations and guidelines are also the quintessence of best practices for data sharing. Given this, the 'Discussion' section would benefit from placing more emphasis on the standardizing work done on the subject matter of the report. This could be achieved by at the very least elaborating on the direct and indirect effects of the work and the delivered results as a condition of the efficient application of best practices.

As already highlighted, the focus on the procedure of creating the recommendations and guidelines is of the utmost relevance both for responding to the current pandemic as well as in the sense of creating a living document as a blueprint for dealing with future global challenges. Accordingly, the description of the consultation process is incredibly important ("expert-informed", "community reviewed"). It would therefore be useful to provide more details on the consultation participants, the exact frequency, methods and modi of consultation and how the results of the consultation have been taken into account.

Secondly, the report can be characterized – suitably – as an open and responsive document. Nevertheless, in the text of the article, it is not yet clearly described whether there might still be a chance for further consultation. If such a chance still exists, which would be understandable given the fast-paced development of the areas focused on in the report since its publication as well as the valuable guidance provided by other scientific communities in the last months, the proposed methods of openness of the report itself should be shortly elaborated upon. Sounding out approaches to developing a "learning" and "living" document would be highly appreciated by various affected communities, even where this requires additional effort from those involved in its creation.

Thirdly, consultation with additional international organizations such as sub-organizations of the UN (UNESCO, WHO) would be beneficial, should this not yet have occurred, as both sub-organizations of the UN are frequently cited in the report. If consultation has already been conducted, these organizations should be mentioned in the article (besides the OECD).

The fifth set of suggestions relates to the additional tools cited in the review.

It would be beneficial to clarify in the article for whom the decision-making tool is intended. All addressees need to make decisions in their respective contexts and impact areas. The tool would thus be perfect for demonstrating the implementation and application of the recommendations and guidelines through use-case models, and I highly recommend its use as such. It would allow the various addressees to see for themselves the applicability and helpfulness of the report's guidance in real-life scenarios.

Furthermore, there is the question of the relationship between the decision-making tool and the DS Wizard Navigation Tool – are these the same, and, if not, what is the connection between them? (Unfortunately, I could not register to try the tool.)

The question also arises as whom the mind map is intended for. Clarification in the article would

be helpful as to whether it relates to a specific circle of addressees or instead presents the proceeding of the working group / drafters / contributors. Furthermore, although this is a technical issue, the mind map does not fit on a standard laptop monitor. While it obviously still needs to be captured at a glance, the textual descriptions can be blended in as notes and zoom and filter functions do exist, an additional, more structured version made for reading on smaller devices would be useful.

The infographic is an important tool for communicating the results. While it is appreciated that it presents the essence of the results in an easily understandable and consumable fashion, some minor extensions by just one or two words in the relevant places could better emphasize the main results in relation to the relevant parts of the report presentation. Although not explicitly relevant for the text of the article, the weighing of “ethics vs privacy” is slightly generalized in the infographic. A more nuanced view on ethics might be transmitted with one or two additional adjectives, which would also better communicate the balanced results of the report.

Finally, and in summary, I would like to emphasize – in addition to the breadth of interdisciplinary effort that makes the initiative and the report stand out among other endeavors – two crucial aspects of the relevance of the report that could be better highlighted in the article. First the report underlines “[t]he priority [...] for these guidelines and recommendations”, i.e., “to be widely adopted in order to accelerate solutions to the pandemic”. This wide adoption will be achieved by the application of the guidelines and recommendations, with said application creating their inherent consequence – rules of conduct further crystallizing best practices (cf. comments on the decision-making tool above).

Secondly, the importance of the process of developing the report could be further highlighted, as the process includes the involvement of and consultation with stakeholders and the implementation of their approaches throughout the development procedure of the guidelines and recommendations. The development procedure can – through suitable deliberation – inherently foster the substantive appropriateness of the content. The presentation of the report in the article should justifiably demonstrate this interconnectedness.

Is the rationale for the Open Letter provided in sufficient detail?

Yes

Does the article adequately reference differing views and opinions?

Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Yes

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Partly

Competing Interests: The grant from the Canadian Cancer Genome Collaboratory funding the present paper is related to the Canada Foundation for innovation/EUCAN work conducted in Canada as part of the EUCANCan Horizon 2020 project, as part of which the reviewer is funded by the European Commission (grant agreement No. 825835).

Reviewer Expertise: Law.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 20 November 2020

<https://doi.org/10.21956/wellcomeopenres.18015.r41349>

© 2020 Simperl E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Elena Simperl

King's College London, London, UK

The open letter summarises the results of an RDA working group on COVID-19, which aimed to support worldwide data sharing through interdisciplinary recommendations and guidance. The letter describes the composition of the working group and the process followed to come up with the recommendations. It also lists the main findings for each of the four disciplines of concern as well as some cross-cutting themes such as community participation, law and ethics, and software.

This is a timely initiative, covering a huge, interdisciplinary space. The recommendations discussed in the letter are aligned with established best practices and knowledge in research data sharing. My only concern is that they remain very generic - speaking as a computer scientist with a reasonable knowledge of data sharing, FAIR, open science, citizen science etc, I found it difficult to:

1. Appreciate what is new about the recommendations compared to the previous excellent work done by the science community in this space. Many of the recommendations have been discussed extensively before in other contexts. It would have been useful to discuss what is truly new and different - there is an argument around the scale and pace of data sharing, but it does not cut through the article.
2. As the guidance remains rather generic, it wasn't straightforward for me to understand how it would be used in practice and by whom. The letter mentions a diverse range of stakeholders, which are no doubt relevant in this context, but the next steps following the guidance are not easy to grasp. The paper should improve on this aspect to ensure the substantial amount of work that the working group has completed bears fruit in the real world.

Is the rationale for the Open Letter provided in sufficient detail?

Yes

Does the article adequately reference differing views and opinions?

Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Yes

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computer science.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
