OPEN LETTER

REVISED Fostering global data sharing: highlighting the recommendations of the Research Data Alliance COVID-19 working group [version 2; peer review: 2 approved, 1 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 Diallo[10], Thomas Duflot [11], Leyla Garcia [12], Sandra Gesing[13], Alejandra Gonzalez-Beltran [4], Anupama Gururaj [14], Natalie Harrower [5], Dawei Lin[14], Claudia Medeiros [15], Eva Méndez [16], Natalie Meyers [17], Daniel Mietchen [18], Rajini Nagrani [19], Gustav Nilsonne [20], Simon Parker[21], Brian Pickering [22], Amy Pienta[23], Panayiota Polydoratou [24], Fotis Psomopoulos [25], Stephanie Rennes [26], 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

[1]Environment and Climate Change Canada, 351 boul. St-Joseph, Gatineau, Quebec, K1A 0H3, Canada
[2]Centre of Genomics and Policy, McGill University, 740, avenue Dr. Penfield, suite 5200, Montreal, Quebec, Canada
[3]Institute for Science, Innovation and Society, University of Oxford, 64 Banbury Road, Oxford, OX2 6PN, UK
[4]UKRI-STFC Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK
[5]Digital Repository of Ireland, Royal Irish Academy, 19 Dawson St, Dublin 2, D02 HH58, Ireland
[6]CNRS, Inserm and University Toulouse III Paul Sabatier, Toulouse, France
[7]Native 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
[8]Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephessiou, Athens, 11527, Greece
[9]HTW Berlin University of Applied Science, Wilhelminenhofstraße 75A, Berlin, 12459, Germany
[10]BPH INSERM1219 & LaBRI, Univ. Bordeaux, 146 rue Léo Saignat, F-33000, Bordeaux, France
[11]Normandie Univ, UNIROUEN, CHU Rouen, Department of Clinical Research, Rouen University Hospital, 1 Rue de Germont, Rouen Cedex, 76031, France
[12]ZB MED Information Centre for Life Sciences, Gleueler Str 60, Cologne, 50931, Germany
[13]University of Notre Dame Center for Research Computing, 814 Flanner Hall, Notre Dame, IN, 46556, USA
[14]National Institute of Allergy and Infectious Diseases, National Institutes of Health, 5601 Fishers Lane, Rockville, MD, 20852, USA
[15]Institute of Computing, University of Campinas, Av Albert Einstein 1251, Campinas, São Paulo, 13082-853, Brazil
[16]Universidad Carlos III de Madrid, C/ Madrid, 128, Getafe (Madrid), 28903, Spain
[17]250D Navari Center for Digital Scholarship, Hesburgh Library, University of Notre Dame, Notre Dame, IN, 46556, USA
[18]School of Data Science, University of Virginia, P.O. Box 400249, Charlottesville, VA, 22904, USA
[19]Leibniz Institute for Prevention Research and Epidemiology, Achterstrasse 30, Bremen, 28359, Germany
[20]Karolinska Institutet & Swedish National Data Service, Nobels väg 9, Stockholm, 17177, Sweden
[21]Cancer 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

30Indepedent 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

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 clinicians, researchers, policy- and decision-makers, funders, publishers, public health experts, disaster preparedness and response experts, infrastructure providers from the perspective of different domains (Clinical Medicine, Omics, Epidemiology, Social Sciences, Community Participation, Indigenous Peoples, Research Software, Legal and Ethical Considerations), and other potential users. 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

Open Peer Review

Reviewer Status ? ✔ ✔

	Invited Reviewers		
	1	**2**	**3**
version 2 (revision) 26 May 2021		✔ report	
version 1 09 Nov 2020	? report	? report	✔ report

1. **Elena Simperl**, King's College London, London, UK

2. **Fruzsina Molnar-Gabor**, Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany

3. **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.

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

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, Writing – Review & Editing; **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-Palone MR**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Uhlmansiek M**: Project Administration, Resources, Writing – Review & Editing;

Amendments from Version 1

Abstract -
This has been updated to clarify interested parties.

Introduction -
Paragraph two has been updated to discuss the methodology in developing the document; a new paragraph has been introduced after that to justify cross-cutting themes. The paragraph on outputs in addition to the document has been clarified.

The last paragraph has been rewritten and split to discuss how stakeholders have been addressed and the unique nature of the document.

Recommendations section retitled to Subgroup Recommendations
The preamble has been extended to provide more of an introduction to the content of the section.

Clinical guidelines
Two additional sentences have been added.

Epidemiology guidelines
The last sentence has been updated to describe outputs from the RDA COVID-19 Epidemiology sub-group.

Social Science guidelines
An additional sentence has been added.

Community participation guidelines
An additional sentence at end of section.

Legal and Ethical guidelines
The paragraph has been entirely rewritten.

Overarching recommendations/Foundational elements retitled to Overarching recommendations
An additional sentence has been added at the end of section.

Discussion
Paragraph 4 has been updated to describe in more detail how the development of the RDA working group was run.

Two additional paragraphs on additional activities that have been carried out on the basis of the document.

Conclusions
There is substantial rewrite of paragraph three.

References
Additional references added.

Any further responses from the reviewers can be found at the end of the article

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 have 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 (Callaghan, 2020). The working group was divided into four research areas (Clinical, Omics, Epidemiology, Social Sciences) with four cross-cutting themes (Community Participation, Indigenous Data, Research Software, Legal and Ethical Considerations). Despite scheduling challenges across multiple time zones, weekly writing sprints were organised starting with a subgroup discussion, effectively a scrum deciding what issues should be addressed with all subgroups feeding back to one another about progress and specific topics and recommendations. After an overall editorial team had overseen the final editing of all the recommendations, the final draft was released to the wider community for review and comment. Individual working groups were tasked with addressing any issues raised.

The cross-cutting themes were selected as they have impact on all four research areas. Community Participation highlights the work done by communities who are collecting, curating, and sharing data with the goal of improving research outputs and public knowledge. Each of the research areas are dependent on these activities. Indigenous Peoples are acutely impacted by the negative social, economic, environmental and health outcomes of COVID-19 (UN Special Rapporteur on the rights of Indigenous Peoples, 2020) and hence researchers must be aware of their particular needs. Legal and Ethical Considerations are necessary to inform researchers, practitioners and policymakers on how to deal with these aspects of pandemic response (European Group on Ethics in Science and New Technologies, 2020; UNESCO, 2020; WHO, 2007), balancing principles of openness with concerns related to human rights and dignity (Council of Europe, 2020). Research Software is key in conducting research (Nangia & Katz, 2017). Developing and publishing research software (discussed in the RDA document) enables reproducibility, rapid development, and correction of the software and hence impacts all the above four research areas.

The objective of this RDA working group was to provide data sharing recommendations for researchers, clinicians, policymakers, funders, publishers, and providers of infrastructure concerning the most important challenges encountered during 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 data sharing 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. In addition, there was the possibility for sub-sections to provide a link to additional supporting output in the form of discussion papers or preprints (RDA COVID-19 Epidemiology WG, 2020).

Prefaced by an executive summary, the Recommendations and Guidelines provide an essential reference text for the above stakeholders. Each section provides high level recommendations to policymakers and funders, followed by more granular guidelines for the other stakeholders. It is also supported by an extensive curated bibliography publicly accessible online Zotero Library (RDA COVID-19 WG, 2021) and which continues to be updated to support ongoing work following publication of the Recommendations and Guidelines. An infographic was created to provide an overview and highlight key areas. Two prototype tools are under development: (a) the DS Wizard, that allows readers to pull out content to create an abridged version focused on their specific interests; and, (b) a mind map to assist readers in exploring the document. These prototype tools under development are available on the Value of RDA for COVID-19 webpage. The comprehensive recommendations and guidelines and related navigation tools facilitate uptake by all stakeholders (including the public), who wish to access 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 interoperability and data exchange between stakeholders. It highlights the advances and procedures in different disciplines, but crucially also draws attention to commonalities across disciplines, fostering interdisciplinary action, understanding of the disciplines that stakeholders are not part of, and future collaboration. Although recommendations are most frequently aimed at stakeholders such as researchers, they also include items of relevance to policy and decision-makers around governance and data protection, legislation, and funders in encouraging appropriate planning from the outset for managed data sharing. Explicit guidance per stakeholder is also provided in the navigation tools accompanying the recommendations.

The RDA is well positioned to develop such guidance due to its grassroots, participative tradition of interdisciplinary self-motivated dialogue and solutions-based outputs. Indeed, the sheer number of experts globally prepared to commit time during the very tight timeline imposed by the public health emergency makes the recommendations a model example for such all-encompassing collaboration. Beyond definitive guidance, the output may also serves to encourage further discussion and action. It is a significant body of work highlighting a common understanding and motivation to share knowledge across the research community.

Subgroup recommendations
In this section, we provide a brief motivation for each sub-group, the problems identified, and a summary of key recommendations

per group as well as overarching guidance. As mentioned briefly above, there were four discipline-specific subgroups (Clinical to Social Sciences guidelines, below), focusing on the specific challenges of their domain. However, an important feature of the recommendations were the four complementary working groups (Community participation to Legal and Ethics guidelines, below) described as 'cross-cutting' because of their relevance across the four discipline-specific subgroups, ultimately they provide a much needed domain agnostic-perspective on the one hand relevant to all groups but also typical of the need during such public emergencies to review and understand global issues which inform and unite different aspects of research.

One common motivator for the recommendations is to find mechanisms that allow data sharing whilst maintaining appropriate governance. In consequence, the Social Science recommendations, for example, emphasise both normative and technical interoperability, as well as harmonized access to curated data repositories, whilst encouraging a suitable balance between the rights of those providing data and the potential benefits to the community. This topic is addressed in detail by the Community participation subgroup especially from the perspectives of app development for community-generated data (for tracking and contact tracing), collaborative data collection and stewardship.

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. The focus was on data in clinical trials and clinical care outside clinical trials. Standards on immunological, imaging, and other healthcare data are identified. 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 underlying the research available alongside the 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 policymaking based weak or fraudulent studies, which in turn causes distrust in science (The Editors of the Lancet Group, 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 RDA Omics sub-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 COVID-19 epidemiology is that data and models are not comparable or interoperable, and they are frequently incomplete, provisional, and subject to correction under changing conditions, making their use and reuse for timely epidemiological analysis extremely challenging. The principal guidelines for researchers are to ensure that the data models are inclusive of not only clinical data, disease milestones, indicators and reporting data, but also contact tracing and personal risk factors. Our recommendations for policymakers are to incentivise the publication of situational data, analytical models, scientific findings, and reports used in decision making based on common standards so that they are reusable by epidemiologists. The supporting output (RDA COVID-19 Epidemiology WG, 2020) expand upon six focus areas (data sources, instruments, privacy, epidemiological data model, computable framework, and an epi-stack framework) to progressively develop a data driven global vision for managing novel biological threats such as COVID-19.

Social sciences guidelines

The social sciences recommendations seek to ensure that social science data are 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. Given the cross-disciplinary and pervasive nature of social science data especially during socially disruptive events such as the coronavirus pandemic, the recommendations focus more specifically on the normative and technical aspects of data interoperability and exchange. The subgroup recommendations therefore 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 while ensuring the protection of human subjects and that confidential data are kept secure.

Community participation guidelines

Community participation guidelines were created with the aim of bridging stakeholder involvement and ensuring that inputs from researchers, citizen scientists, developers and device makers are streamlined, while perspectives from patients, policymakers and the public at large are also considered. Linking communities and supporting communication is essential for coordination and avoiding 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 and provide a set of guidelines designed to ensure an approach based on best participatory practices. Furthermore, the Community Participation guidelines also aim to enable citizen scientists undertaking research to contribute to a common body of knowledge, and to encourage public and patient involvement (PPI) throughout the data management lifecycle from research question to final data sharing and usage especially in the emergency contexts that require their direct and active engagement.

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 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. 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 improvement of 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 cover 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. Especially for health-related data, this frequently leads to tension between the interests of the individual and of society in general. This section therefore makes recommendations to help ensure best practices are respected in using COVID-19 data across jurisdictions and institutions. The guidelines draw attention to approaches which may be of particular interest to policymakers and regulators. These recommendations include a synthesis of foundational principles of data privacy in law and ethics, a description of organisational data governance practices, and sources of legal and ethical obligations applicable to researchers performing studies during COVID-19, including biomedical and social science research ethics guidance. Data governance is considered throughout the data lifecycle in the spirit of community engagement and benefit sharing and leads on to a discussion of the distinct consent standards applicable to clinical care, research ethics, and data privacy law. Practical pointers help researchers identify the most appropriate actor at their institution to guide them in adhering to local legal and ethical requirements, while best practices for data de-identification and anonymisation, as well as data and software IP licensing are described. In many instances, sharing data with external researchers or transferring data to a third country can engage competing legal responsibilities and create legal ambiguities or impose conflicting obligations on researchers. Since multiple regulatory regimes may apply simultaneously to a single instance of data use or data sharing, therefore, the recommendations emphasize the need for research institutions and regulatory bodies to participate in the clarification and interpretation of the ethical and legal principles for data use, data sharing, and international data transfer.

Overarching recommendations

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 technically 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 data 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. Ethical and reproducible data were also emphasized, leading to the the concept of FAIRER data.

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 external sharing and secondary use 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 are 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 among researchers from multiple disciplines and the transfer of data across 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 and 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 and drafting of a consensus document. The priority is to encourage wide adoption of these guidelines and recommendations in order to help accelerate successful solutions to the pandemic.

Instead of a silo-based approach, the document identifies the commonalities in data management across different research areas and themes. It is the result of a standardised common approach in how the different sections were drafted, structured and reviewed. 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 can provide valuable insights and stimulate a creative process. The added value of such overarching collaboration is a key takeaway from this process that may also enrich similar efforts.

The document was developed with a comparatively light level of moderation and emerged on a very rapid timeframe of 10 weeks, including the release of five drafts posted for open feedback and comments on a weekly basis. Writing coordination focused on ensuring the flow of information, so sub-groups, moderators and chairs met regularly. There was a weekly public webinar, as well weekly Co-Chairs meetings, and weekly coordination sessions for Chairs and Moderators. In addition, Subgroups, led by co-moderators, determined their respective meeting frequency and manner of working, ranging from one to three or more meetings per week depending on the group. Subgroups were responsible for reviewing and resolving any comments received on the previous week's work. Small teams were set up for visualisation of recommendations, and for managing references across all sub-groups. 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 lightweight structure was enabled through relatively simple tools, namely Google Docs, Zotero and videoconference calls. The final publication is designed as a reference text, where users 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 demonstrated 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 are already in place, so we need to foster this through continued coordination, harmonisation, and decision making. Engagement within different stakeholder groups has continued, representatives from the European Commission developed a

factsheet and informed partner organisations including The Coalition for Epidemic Preparedness Innovations, WHO, Wellcome, and the Gates Foundation.

This work has led to a number of follow-up activities within the RDA, including the creation of a new Community of Practice for Infectious Disease Data. There were four follow-up sessions at the 16th RDA Plenary in November 2020, and another three are planned for the 17th Plenary in April 2021. There have also been many spinoff activities. One article has been published in a peer-reviewed journal (Rodriguez-Lonebear *et al.*, 2020), one is currently under review (Carroll *et al.*, 2021, submitted to Frontiers in Medical Sociology), and one has been conditionally accepted (Pickering *et al.*, in press, submitted to Open Research Europe), in addition to the present paper. There are a further five papers that have been published as preprints (Austin *et al.*, 2020; Hallinan *et al.*, 2020; Sauermann *et al.*, 2020; Schmidt *et al.*, 2020; and Tonnang *et al.*, 2020), three that are available as discussion papers (Greenfield *et al.*, 2020a; Greenfield *et al.*, 2020b; Greenfield *et al.*, 2020c), and one that is available as an RDA supporting output (Harrower & Dillo, 2020). There have been 10 conference presentations and seven webinars. Finally, the results of this work have supported a number of research grant proposals including two that have been funded relating to COVID-19 and Indigenous communities (NIEHS, 2020).

These spinoff activities demonstrate a clear direct impact of this initiative for both the advancement of research and the leveraging of "expert-informed" and "community reviewed" resources. The RDA is committed to sharing and improving the approach as an example of good practice, offering its structure, processes, and support as a framework for similar efforts.

Conclusions

The RDA COVID-19 Recommendations and Guidelines on Data Sharing (RDA COVID-19 WG, 2020) highlights the importance of data sharing and secondary data use 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 policymakers, 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.

The present paper has focused on the above document and the WG. An analysis of other related community activities is beyond the scope 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 multidisciplinary, bottom-up and geographically diverse community, to be able to rapidly respond to acute global challenges such as 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. The guidelines can be an important resource for the Organisation for Economic Co-operation and Development (OECD), the Bill and Melinda Gates Foundation, Wellcome, the WHO, the Global Research Collaboration for Infectious Diseases Preparedness (GloPID-R), the European and Developing Countries Clinical Trials Partnership (EDCTP) and the Innovative Medicines Initiative (IMI).

The experience of writing the guidelines demonstrates 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 be it 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. The question remains of how similar large scale, multidisciplinary challenges might 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 reflect a large diverse community. The process of forming the collaboration and developing the guidelines was also studied as an object of social science research.

Data availability
No data are associated with this article.

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.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Austin CC, Widyastuti A, Jundi NE, *et al.*: **COVID-19 Surveillance Data and Models: Review and Analysis, Part 1.** 2020.
Publisher Full Text

Callaghan S: **Data Sharing in a Time of Pandemic.** *Patterns (N Y).* 2020; **1**(5): 100086.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Carroll SR, Akee R, Chung P, *et al.*: **Indigenous Peoples' Data During COVID-19: From External to Internal.** *Front Sociol.* 2021; **6**: 617895.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Clément-Fontaine M, Di Cosmo R, Guerry B, *et al.*: **Encouraging a Wider Usage of Software Derived from Research.** Committee for Open Science's Free Software and Open Source Project Group. 2019.
Reference Source

Council of Europe: **The impact of the COVID-19 pandemic on human rights and the rule of law.** 2020.
Reference Source

European Group on Ethics in Science and New Technologies: **Statement on European Solidarity and the Protection of Fundamental Rights in the COVID-19 Pandemic**. 2020.
Reference Source

Greenfield J, Nagrani R, Sears M, *et al.*: **A Full Spectrum View of the COVID-19 data domain: An Epidemiological Data Model.** In *COVID-19 Data sharing in epidemiology, version 0.06b.* Research Data Alliance RDA-COVID19-Epidemiology WG. 2020a.
Reference Source

Greenfield J, Sears M, Nagrani R, *et al.*: **Common Data Models and Full Spectrum Epidemiology: Epi-STACK architecture for COVID-19 epidemiology datasets.** In *COVID-19 Data sharing in epidemiology, version 0.06b.* Research Data Alliance RDA-COVID19-Epidemiology WG. 2020b.
Reference Source

Greenfield J, Tonnang HEZ, Mazzaferro G, *et al.*: **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 RDA-COVID19 WG, *COVID-19 Data sharing in epidemiology, version 0.06b.* Research Data Alliance RDA-COVID19-Epidemiology WG. 2020c.
Reference Source

Hallinan D, Bernier A, Cambon-Thomsen A, *et al.*: **International Transfers of Health Research Data Following Schrems II: A Problem in Need of a Solution.** 2020.
Publisher Full Text

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

Nangia U, Katz DS: **Understanding Software in Research: Initial Results from Examining Nature and a Call for Collaboration.** *2017 IEEE 13th International Conference on E-Science (e-Science).* 2017; 486–487.
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**

NIEHS: **Investigating linkages between arsenic exposure, diabetes, and COVID-19 infections and risks on the Navajo Nation.** September 2020 – August 2022. PI: Stephanie Russo Carroll. Multi-PI J. Ingram. Funded by the National Institute for Environmental Health Sciences ($154,199). 2020.
Reference Source

Pickering B, Biro T, Austin CC, *et al.*: **Radical collaboration during a global health emergency: Development of the RDA COVID-19 data sharing recommendations and guidelines.** *Open Research Europe.* (in press).

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 COVID-19 Epidemiology WG: **Sharing COVID-19 Epidemiology Data.** (version 1.0). Research Data Alliance. 2020.
Publisher Full Text

RDA COVID-19 Working Group: **Recommendations and Guidelines on Data Sharing.** (version 1.0). Research Data Alliance working group. 2020.
Publisher Full Text

RDA COVID-19 WG: **The RDA COVID-19 Zotero Library.** (version 1.0 released on June 30, 2020. Version 2.0 released on March 7, 2021). Research Data Alliance working group. 2021.
Publisher Full Text

Rodriguez-Lonebear D, Barceló NE, Akee R, *et al.*: **American Indian Reservations and COVID-19: Correlates of Early Infection Rates in the Pandemic.** *J Public Health Manag Pract.* 2020; **26**(4): 371–377.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sauermann S, Kanjala C, Templ M, *et al.*: **Preservation of individuals' privacy in shared COVID-19 related data.** 2020.
Publisher Full Text

Schmidt CO, Nagrani R, Stange C, *et al.*: **COVID-19 questionnaires, surveys, and item-banks: Overview of clinical- and population-based instruments.** 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**

Tonnang HEZ, Greenfield J, Mazzaferro G, *et al.*: **COVID-19 Emergency public health and economic measures causal loops: A computable framework.** 2020.
Publisher Full Text

UNESCO: **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.
Reference Source

WHO: **Ethical considerations in developing a public health response to pandemic influenza.** (WHO/CDS/EPR/GIP/2007.2). World Health Organization. 2007.
Reference Source

Wilkinson MD, Dumontier M, Aalbersberg IJJ, *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 2

Reviewer Report 07 July 2021

https://doi.org/10.21956/wellcomeopenres.18520.r44088

✓ **Fruzsina Molnar-Gabor**

Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany

I have assessed the recent revisions and think they are appropriate.
I am looking forward to reading the announced further publication that will analyse how this community worked together at speed to deliver the final document.

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).

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.

Version 1

Reviewer Report 13 January 2021

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

✓ **Maria Luiza Machado Campos** (iD)

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.

Author Response 02 Apr 2021

Hugh Shanahan, Royal Holloway, University of London, Bedford Building, Egham, UK

We thank the referee for her comments and have addressed the points made. We list below the comments made and our response to each.

"Recommendations section: could benefit from better balancing the content of the different subgroups or areas. Simpler section 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."
The relevance of the different cross-cutting themes, including Community participation, has been explained in more detail in section 1, paragraph 3. Furthermore the Community Participation overview has been additionally clarified in section 2.5. The section on clinical guidelines has also been updated.

"Discussion section: 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. Maybe the joint effort Data Together, involving RDA, CODATA, GO FAIR and EOSC could have been mentioned."
In section 6 (discussion) paragraph 5 the impact of other initiatives is discussed.

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

Competing Interests: No competing interests were disclosed.

Reviewer Report 07 January 2021

?   **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.

> Author Response 02 Apr 2021
> **Hugh Shanahan**, Royal Holloway, University of London, Bedford Building, Egham, UK
>
> We thank the referee for her comments and have addressed the points made. We list below the comments made and our response to each.
>
> *"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."*
> **A more detailed summary and presentation of the four cross-cutting themes have been made (section 1, paragraph 3) and an explanation why it is relevant in an emergency situation is explained at the start of section 2.**
>
> *"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."*
> **The inconsistencies in the structure of the document have been remedied - in particular "Subgroup Recommendations" and "Overarching Recommendations" have been separated into two distinct sections.**
>
> *"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."

These delineations are more concerned with the provenance of the recommendations than the specific audience to which they are directed. Mirroring the structure of the subgroups in the overarching themes section and in the subsections may prove impracticable, as the guidance is drafted in a holistic fashion so as to increase its multidisciplinary appeal. With this in mind, we have not updated the paper with respect to this.

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

This inconsistency has been addressed with clinicians added in the abstract.

"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."

Those terms have been harmonized in section 2.8 (Legal and Ethics Guidelines). Specifically the following nomenclature has been adopted: Data sharing is retained as the generic term used to refer to the exchange of data among different groups. The term data reuse is used to connote circumstances in which the secondary use of data raises special technical, operational, or ethical considerations. The term international data transfer is used to refer to instances in which the sharing of data across national boundaries raises particular legal issues. In this section data use, data sharing, and international data transfer are differentiated.

"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."

The use of such terms has been rationalised in the text, with reference to them removed in section 1, paragraph 6 and and section 2.6.

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

Technical and legal/normative interoperability have been labelled as such. (Section 2 paragraph 2; section 2.4)

"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."

The guidelines and recommendations in the area of social sciences have been emphasized (Section 2.4).

"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)"
The legal challenges have been better emphasized - specifically a sentence has been added to the ethical and legal section highlighting the instrumental role of research institutions and of regulatory bodies in helping researchers navigate overlapping ethical and legal regimes (end of section 2.8).

"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 section on Legal and ethical guidelines have been reworked such that different sentences are used to highlight the distinct elements of the guidelines directed at the scientific community, and those addressed to policymakers. To ensure that the section adopts the same structure as those of the other subgroups, the section has not been divided into separate paragraphs.

"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."
A new paragraph (section 4 paragraph 6) has been added to address follow up and spin off activities based on the initial work. The discussion section have been emphasized on the standardizing work done elaborating on the direct and indirect effects of the work and the delivered results (section 5, paragraphs 3, 5 and 6)

"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."
The description of the consultation process has been detailed in section 1, paragraph 2 and section 5 (discussion) paragraph 4.

"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."
As noted above a new section has been added before the Discussion section on further activities based on the initial work with additional references across the document to highlight uptake e.g. section 5, paragraph 6.

"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)."
We have mentioned consultation with additional international organizations in section 1, paragraph 7 and section 5 (discussion) paragraph 4.

"*It would be beneficial to clarify in the article for whom the decision-making tool is intended. All addresses 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 addresses 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.)"
We have clarified the status of the tools being developed in section 1, paragraph 6.

"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."
We have clarified the role of the mindmap (section 1, paragraph 6). We acknowledge that the mindmap is not optimized for reading on laptops and smaller devices. We also note the text specifies that the mindmap is still in development.

"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."*
> **The infographic is being updated as suggested.**
>
> *"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)."*
> **Crucial aspects of the relevance of the report have been emphasized in section 1, paragraphs 7 and 8.**
>
> *"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."*
> **This will be discussed at length in a companion another paper that will analyse how this community worked together at speed to deliver the final document. Hence we believe it appropriate to not preempt the findings of that paper.**
>
> *Competing Interests:* No competing interests were disclosed.

Reviewer Report 20 November 2020

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

?

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.

Author Response 02 Apr 2021

Hugh Shanahan, Royal Holloway, University of London, Bedford Building, Egham, UK

We thank the referee for her comments and have addressed the points made. We list below the changes made to the text to address the comments.

"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:
-   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."
We now discuss what is truly new and different compared with the work done by the science community in this space (section 1, paragraphs 7 and 8). Aspects related to the exceptional scale and need from timely response have been additionally addressed (section 2, paragraph 1 and 2; section 6 paragraph 5).

" 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."
We have added information on the best way to use the guidance in the real world (section 1, paragraphs 6, 7 and 8).

Competing Interests: No competing interests were disclosed.