

It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too!

Gabriela Karolina Hajduk,¹ Nina E Jamieson,¹ Bonny L Baker,¹ Ole F Olesen,² Trudie Lang¹

To cite: Hajduk GK, Jamieson NE, Baker BL, *et al.* It is not enough that we require data to be shared; we have to make sharing easy, feasible and accessible too! *BMJ Global Health* 2019;**4**:e001550. doi:10.1136/bmjgh-2019-001550

Handling editor Seye Abimbola

Received 6 March 2019

Revised 30 May 2019

Accepted 8 June 2019

Over the past several years, we have seen a movement towards a more open way of conducting science, with recommendations that ought to lead to reproducible methods, analyses and results, as well as reusable data. Data sharing is widely encouraged and its importance has been noted in the context of health data, including clinical trials.^{1 2} It is now a standard requirement by publishers, research institutions and regulatory agencies to share data. Many types of health data are increasingly viewed as global public goods that should be made available to the wider scientific community without unnecessary delays, ensuring important findings can be extracted as soon as possible.^{3 4} Major funders such as the European Commission, National Institutes of Health, Wellcome Trust, Bill & Melinda Gates Foundation and The European and Developing Countries Clinical Trials Partnership (EDCTP; www.edctp.org) are imposing contractual obligations on their grantees to share their data for free and, ideally, without imposing unnecessary barriers on data accessibility (ie, open access or appropriate controlled access⁵). In the current landscape of global health research, data sharing, including sharing of clinical data collected during routine patient care, clinical data collected by clinical trials, as well as metadata, has therefore become a simple necessity.

NOT AS EASY AS IT SOUNDS

Despite these requirements, sharing of health data in a meaningful manner is neither straightforward nor commonplace. We suggest that this is at least partially due to the lack of clear standards and established guidelines explaining where, when and how to share data. We carried out a gap analysis in order to assess the needs of researchers, as well as the resources and training available to

Summary box

- ▶ The sharing of health data, including clinical trial data, is required more and more often by research publishers, regulatory agencies, ethics committees and funding bodies.
- ▶ Despite these requirements, there are currently no clear standards and guidelines of how, where and when researchers should share their data.
- ▶ The confusion among researchers regarding issues related to data sharing has led funders such as The European and Developing Countries Clinical Trials Partnership (EDCTP) to devise initiatives that will provide their grantees, and the wider scientific community within the field of global health research, with clear guidance and a range of tools to facilitate the data sharing process.
- ▶ In an effort to support and facilitate data sharing, the EDCTP is working in collaboration with The Global Health Network to assess whether a cross-cutting knowledge hub around data sharing would help researchers find the optimum repository and to gather their data in a form that is ready for sharing.

them. Our approach was threefold. (1) We used specialist web browsing software to carry out a comprehensive audit of online training courses, learning materials and educational videos related to data sharing in health research by querying Bing, Exalead, Google, Yahoo and YouTube. (2) We conducted a workshop on data sharing and obtained feedback from the attendees regarding their training needs in this field. (The workshop was organised by The Global Health Network in collaboration with the Infectious Diseases Data Observatory and carried out during the EDCTP Ninth Forum (17–21 September 2018) in Lisbon, Portugal.) (3) We investigated repository availability and their characteristics in order to develop a tool that will guide researchers to repositories appropriate for their datasets.



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK

²International Cooperation Europe, European and Developing Countries Clinical Trials Partnership, The Hague, The Netherlands

Correspondence to

Trudie Lang;
trudie.lang@ndm.ox.ac.uk

Data sharing is complicated and costly in terms of time, effort, expertise and resources. There are of course other obstacles, including concerns about data sensitivity and patient privacy, as well as the technical aspects of data processing before the data can be shared.^{2-4 6 7}

These challenges hold true across multiple contexts, for example, not just among researchers but also the public⁸, in high-income settings as well as low-income and middle-income settings. Overall inequality in health data can be linked to poverty,⁹ and similarly data sharing may be particularly challenging for researchers in low-income and middle-income countries (LMICs).¹⁰ For instance, inequities exist between high-income countries (HICs) and LMICs when it comes to data ownership and reuse.¹¹ One of the main concerns of primary researchers is that while they spend time and effort collecting and sharing data, secondary researchers will focus on reusing these data and reaping the benefits, potentially without proper acknowledgment of the primary researchers, and without having contributed to the costs of data generation and processing.^{12 13} Furthermore, LMIC researchers may not even be able to access outputs of such secondary analyses produced using their own data, particularly if these are published behind a paywall in a HIC, and thus their communities will not be able to benefit from the advancements. Moreover, LMIC researchers will likely be responsible for the necessary community engagement and any ethical concerns of their study participants relating to informed consent and data sharing. On top all of these challenges, LMICs also face problems of limited resources and difficulties in accessing the training necessary to build research capacity for data management, processing, analysis and sharing.^{11 14-17}

The nature of working with data is changing at an unprecedented rate due to advancements in technology and analytics techniques.^{18 19} Therefore, it is not sufficient to simply require data to be shared, without providing guidance and assistance with the process, especially if the objective is to share the data in a responsible and useful way. Yet, we struggled to find organisations that provide tools and resources necessary to fulfil their requirements of data sharing. Furthermore, the situation is not helped by the lack of follow-up from the organisations requiring that data are shared. Given that there are few incentives and multiple barriers to data sharing, regardless of whether these incentives and barriers are actual or perceived, as well as lack of support and, ultimately, of consequences, perhaps it is not surprising that data sharing has not been taken up more quickly.

WHAT CAN WE DO?

Some of the existing initiatives supporting data sharing include platforms providing advice, such as the Digital Curation Centre (<http://www.dcc.ac.uk>), the Research Data Alliance (<https://rd-alliance.org>) and Chatham House's guide to sharing health surveillance data (<https://datasharing.chathamhouse.org>), repositories where data

can be archived (with re3data.org collating multiple repositories), consortia working on standards supporting interoperability between different systems (eg, the Clinical Data Interchange Standards Consortium (<https://www.cdisc.org>)), groups developing tools for specific diseases (eg, Malaria Toolkit (Infectious Diseases Data Observatory; <https://www.wwarn.org/tools-resources/malaria-clinical-trials-toolkit>), Ebola Data Tools (ISARIC; <https://isarc.tghn.org/protocols/ebola-data-tools/>), Zika Research Tools (ISARIC, PREPARE Europe, and partners; <https://zika-infection.tghn.org/research-tools-and-resources>)) and trial registries facilitating discovery of the data sets such as ClinicalTrials.gov, ISRCTN (<http://www.isrctn.com>) and the EU Clinical Trials Register (<https://www.clinicaltrialsregister.eu>).

We believe that to address the conceptual difficulties, as well as the legal and ethical concerns, clear and concise information explaining the terminology, funder requirements and policies and the core components of the process of data sharing ought to be easily accessible in a central knowledge hub that is relevant for various health-related data and for a range of study types (from observational to clinical trials), regions and organisations. Bringing the information together has the clear advantage of saving the researcher and/or data manager's time that would otherwise be spent on searching through multiple guides/websites/protocols. Duplication of content should be avoided where practical, for instance, through providing an overview of the issues and signposting to further, more detailed resources. Data sharing should also be put in context—it should be considered throughout the life of the project rather than treated as an afterthought. The resources should reflect this approach.

Capacity development in terms of the technical skills necessary to process, analyse and share data is vital²⁰ and could be addressed through a blend of face-to-face training and complementary online resources. While practical workshops are an impactful way to teach technical skills, they are also expensive to run—and the cost may be prohibitive, particularly in resource-poor settings, limiting attendance to those who can afford to travel to such workshops. Learning materials—such as articles, recordings of seminars, handbooks, and online courses—available without an institutional affiliation and without a fee—would help to bridge the gap and to ensure that everyone who needs to develop these skills is supported in doing so. Providing clear guidance and a variety of resources in one easily identifiable place that can be referred to as needed, should go some way towards addressing the concerns about the time necessary to prepare and share data. In terms of funding for activities related to data sharing, incorporating data sharing into data management plans and funding applications should be supported by providing practical guidance on how to effectively develop such plans.

Funders have an interest in supporting their grantees—and the wider scientific community—with clearer guidance and a range of tools facilitating the process of

Box 1 Resources for practitioners and researchers

- ▶ The Knowledge Hub (<https://EDCTPKnowledgeHub.TGHN.org>) contains a 'Data Sharing Toolkit' covering various aspects of the process of sharing research data—from understanding the different models of access (open/controlled/closed) and options offered by repositories, through preparing the data (e.g. naming conventions, de-identification of sensitive data, using non-proprietary data formats) and the documentation (how to write a README file), to a checklist of files to include with a data submission and an example of the submission process. The Global Health Network and EDCTP are also developing a 'Repository Tool' that will guide researchers through the process of choosing a repository appropriate for their data set.
- ▶ Furthermore, the 'Data Sharing Toolkit' includes a collection of nearly 200 external resources (<https://edctpknowledgehub.tghn.org/data-sharing-toolkit/collated-external-resources/>), including guides, recordings of seminars and comprehensive e-learning courses. Additionally, as data processing is crucial to preparing data sets for sharing, 'The Hub' also covers the basics of data management, which will be developed into a rich 'Data Management Portal' in due course.
- ▶ The objective of 'The Knowledge Hub' is to be a one 'go-to' platform, by providing researchers with start-to-finish guidance on all aspects of working with health data and the eventual data sharing, but also supplying bespoke tools that will make the process easier. All resources available within 'The Hub', as well as anything included in the collection of external resources, are freely accessible to all users regardless of institutional affiliation or funder. It will be important that research teams feedback on the usefulness of this platform.

data sharing. Encouragingly, funders are taking steps to improve the situation by financing projects related to data sharing and, as described above, there are now a few different types of initiatives supporting data sharing. Researchers are also contributing by collating and publishing information in order to facilitate the development of guidelines and principles.^{14 16 21–25} Currently, EDCTP is working in collaboration with the Global Health Network to create a one 'go-to' platform—The Knowledge Hub—that will facilitate all aspects of data sharing in health research.

The Knowledge Hub (EDCTPKnowledgeHub.tghn.org) will provide free and accessible resources, guidance and training on how to manage and share data. This includes resources relevant to all stages of data sharing, from data collection, processing and management, through preparation of metadata and documentation, to guidance on choosing an appropriate repository for data deposition (Box 1). The aim of this Hub is to become a beneficial resource for researchers that can guide and support the process of running a research project, including data sharing. While the focus of EDCTP is on clinical trials, many, if not most, of the Hub's resources should be applicable to working with other health data, and not just limited to clinical trials data. Ongoing feedback from the research community will be essential to refine and validate the

usefulness of this resource and to improving data sharing practices of research teams working with health data.

Acknowledgements We thank the Infectious Diseases Data Observatory members who collaborated with us on the data sharing workshop carried out at the EDCTP Ninth Forum in September 2018, as well as all workshop participants for their time and valuable contributions

Contributors GKH contributed to the conception and writing of the article and revised the article. GKH revised and resubmitted the article with comments from NEJ and BLB. All authors contributed to the review and feedback on the article and approval of the final manuscript.

Funding This work was funded by the European and Developing Countries Clinical Trials Partnership (EDCTP).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Singh KNM, Shetty YC. Data sharing: a viable resource for future. *Perspect Clin Res* 2017;8:63–7.
2. Bull S, Roberts N, Parker M. Views of ethical best practices in sharing individual-level data from medical and public health research: a systematic scoping review. *J Empir Res Hum Res Ethics* 2015;10:225–38.
3. Walport M, Brest P. Sharing research data to improve public health. *The Lancet* 2011;377:537–9.
4. Antman E. Data sharing in research: benefits and risks for clinicians. *BMJ* 2014;348:g237.
5. Sydes MR, Johnson AL, Meredith SK, *et al.* Sharing data from clinical trials: the rationale for a controlled access approach. *Trials* 2015;16.
6. van Panhuis WG, Paul P, Emerson C, *et al.* A systematic review of barriers to data sharing in public health. *BMC Public Health* 2014;14.
7. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA* 2014;311:2479–80.
8. Aitken M, de St Jorre J, Pagliari C, *et al.* Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC Med Ethics* 2016;17.
9. Byass P. The unequal world of health data. *PLoS Med* 2009;6:e1000155.
10. Parker M, Bull SJ, de Vries J, *et al.* Ethical data release in genome-wide association studies in developing countries. *PLoS Med* 2009;6:e1000143.
11. Bezuidenhout LM, Leonelli S, Kelly AH, *et al.* Beyond the digital divide: towards a situated approach to open data. *Sci Public Policy* 2017;44:464–75.
12. Langat P, Pisartchik D, Silva D, *et al.* Is there a duty to share? ethics of sharing research data in the context of public health emergencies. *Public Health Ethics* 2011;4:4–11.
13. Zarin DA. Participant-level data and the new frontier in trial transparency. *N Engl J Med* 2013;369:468–9.
14. Rani M, Buckley BS. Systematic archiving and access to health research data: rationale, current status and way forward. *Bull World Health Organ* 2012;90:932–9.
15. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ* 2010;88:462–6.
16. Whitworth J. Data sharing: reaching consensus. *Bull World Health Organ* 2010;88:468.
17. Rani M, Bekeedam H, Buckley BS, *et al.* Improving health research governance and management in the western Pacific: a WHO expert consultation. *J Evid Based Med* 2011;4:204–13.
18. Lynch C. Big data: how do your data grow? *Nature* 2008;455.
19. Stephens ZD, Lee SY, Faghri F, *et al.* Big data: astronomical or Genomical? *PLoS Biol* 2015;13:e1002195.

20. Wiehe SE, Rosenman MB, Chartash D, *et al.* A Solutions-Based approach to building Data-Sharing partnerships. *EGEMS* 2018;6.
21. Beck EJ, Gill W, De Lay PR. Protecting the confidentiality and security of personal health information in low- and middle-income countries in the era of SDGs and big data. *Glob Health Action* 2016;9.
22. Ienca M, Ferretti A, Hurst S, *et al.* Considerations for ethics review of big data health research: a scoping review. *PLoS One* 2018;13:e0204937.
23. Pisani AR, Kanuri N, Filbin B, *et al.* Protecting user privacy and rights in academic Data-Sharing partnerships: principles from a pilot program at crisis text line. *J Med Internet Res* 2009;21:e11507.
24. Vallance P, Chalmers I. Secure use of individual patient data from clinical trials. *The Lancet* 2013;382:1073–4.
25. Parker M, Bull S. Sharing public health research data: towards the development of ethical data-sharing practice in low- and middle-income settings. *J Empir Res Hum Res Ethics* 2015;10:217–24.