Check for updates

OPINION ARTICLE

Sharing health research data – the role of funders in improving the impact [version 1; referees: awaiting peer review]

Robert F. Terry [1], Katherine Littler[2], Piero L. Olliaro[3]

[1]Research Policy, TDR - The Special Programme for Research and Training in Tropical Diseases, Geneva, 1211, Switzerland
[2]Policy Unit, Wellcome Trust, London, UK
[3]Implementation and Intervention Research, TDR - The Special Programme for Research and Training in Tropical Diseases, Geneva, 1211, Switzerland

v1 **First published:** 15 Oct 2018, **7**:1641 (doi: 10.12688/f1000research.16523.1)
Latest published: 15 Oct 2018, **7**:1641 (doi: 10.12688/f1000research.16523.1)

Abstract

Recent public health emergencies with outbreaks of influenza, Ebola and Zika revealed that the mechanisms for sharing research data are neither being used, or adequate for the purpose, particularly where data needs to be shared rapidly.

A review of research papers, including completed clinical trials related to priority pathogens, found only 31% (98 out of 319 published papers, excluding case studies) provided access to all the data underlying the paper - 65% of these papers give no information on how to find or access the data. Only two clinical trials out of 58 on interventions for WHO priority pathogens provided any link in their registry entry to the background data.

Interviews with researchers revealed a reluctance to share data included a lack of confidence in the utility of the data; an absence of academic-incentives for rapid dissemination that prevents subsequent publication and a disconnect between those who are collecting the data and those who wish to use it quickly. The role of the funders of research needs to change to address this. Funders need to engage early with the researchers and related stakeholders to understand their concerns and work harder to define the more explicitly the benefits to all stakeholders. Secondly, there needs to be a direct benefit to sharing data that is directly relevant to those people that collect and curate the data. Thirdly more work needs to be done to realise the intent of making data sharing resources more equitable, ethical and efficient. Finally, a checklist of the issues that need to be addressed when designing new or revising existing data sharing resources should be created. This checklist would highlight the technical, cultural and ethical issues that need to be considered and point to examples of emerging good practice that can be used to address them.

Keywords

Health research, data sharing, public health emergencies, data standards, data infrastructure, pandemics, curation

Open Peer Review

Referee Status: *AWAITING PEER REVIEW*

Discuss this article

Comments (0)

Corresponding author: Robert F. Terry (terryr@who.int)

Author roles: Terry RF: Conceptualization, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Littler K**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Olliaro PL**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Terry RF *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Terry RF, Littler K and Olliaro PL. **Sharing health research data – the role of funders in improving the impact [version 1; referees: awaiting peer review]** *F1000Research* 2018, **7**:1641 (doi: 10.12688/f1000research.16523.1)

First published: 15 Oct 2018, **7**:1641 (doi: 10.12688/f1000research.16523.1)

Introduction

The benefits of sharing health research data to improve public health have been promoted by international research funders for over a decade but the reality is that the quality and volume of health research data shared, even in emergency situations, remains low[1,2]. This lack of progress seems to reflect a cultural reluctance among researchers to 'give up their data' without any clear benefits returning to them. This concern is heightened among researchers in low resource settings who feel that the requirements to share data, from funders and journals, risk turning them into data exporters unless greater efforts are made to ensure a fairer distribution of benefits. In this paper we draw on our experience of supporting data sharing initiatives and some commissioned research to highlight the barriers to sharing research data and the role research funders might play to improve this situation.

A decade of progress?

In January 2011, a group of research funding organizations published a joint statement on sharing health research data with the aim to promote the efficient use of those data to accelerate improvements in public health. The funders recognized that for data sharing to be most effective, a combination of technical and cultural issues need to be addressed. They framed this approach around three principles which required any data sharing mechanism they supported to be equitable, ethical and efficient (See Wellcome Trust page on sharing research data). (See Box 1).

Box 1. Sharing research data to improve public health the principles in the full joint statement by funders of health research (2010)

Equitable: any approach to the sharing of data should recognise and balance the needs of researchers who generate and use data, other analysts who might want to reuse those data and the communities and funders who expect health benefits to arise from research.

Ethical: all data sharing should protect the privacy of individuals and the dignity of communities, while simultaneously respecting the imperative to improve public health through the most productive use of data.

Efficient: any approach to data sharing should improve the quality and value of research and increase its contribution to improving public health. Approaches should be proportionate and build on existing practice and reduce unnecessary duplication and competition.

Progress on encouraging the sharing of research data has been made over the subsequent decade and it is now common for research grants and journals to require the data underlying a paper or clinical trial to be shared (see PLOS editorial and publishing policies, AllTrials, and NIH data sharing policy.) However, recent public health emergencies with outbreaks of influenza, Ebola and Zika have brought into sharp focus the realization that the mechanisms for sharing data are neither being used or adequate for the purpose, particularly where data needs to be shared rapidly[3–5].

In addition, researchers working in low- and middle-income countries highlight an inequity created by the disadvantage as they see it by the blanket requirements to share their data. Their concern is that sharing their data too soon, or without any restrictions will lead to their data being analysed by others with greater capacity, and no benefit will return to the researchers themselves or the populations they work with. In effect they become data exporters rather than partners. So while there is a lot of emphasis placed on data being Findable, Accessible, Interoperable and Reusable, known as the FAIR approach, many researchers in developing countries fear the reality for them will be far from fair[6–8].

The findings of two surveys and a workshop

To explore this further, we commissioned two surveys to review the governance arrangements and standards within existing data sharing resources. The findings of those studies informed a workshop held in October 2017 with a set of stakeholders representing researchers and funding organizations. All the reports and supporting files are published as open access under a Creative Commons licence and in free-to-access repositories. Readers are strongly encouraged to read that material as the primary source of reference[1,2,9,10].

The first survey – Data Sharing in Public Health Emergencies - focussed on data sharing in public health emergencies concerned with the pathogens named by the World Health Organization as of priority concern because of their epidemic or pandemic potential (see WHO list of Blueprint priority diseases). A review of academic papers published since 2003 relating to these diseases was undertaken and attempts were then made to access the data underlying those publications via the web and through a direct survey of the corresponding authors. Interviews were undertaken with a range of people either conducting or supporting research in these areas and this was supplemented with a review of institutional policies, discussion documents and academic commentaries about standards and norms in data sharing[1,2].

The second survey - Development of International Standards for Online Repositories - was designed to identify which 'standards' were being used in data sharing relating to the neglected diseases. Standards were identified following a review of publically accessible information (via the web or publication) relating to three main areas each with a set of elements describing the standards under those areas[9].

A third report combined the findings of these two surveys and was used to shape thinking at a workshop held in Antwerp, Belgium in October 2017[10].

The workshop brought together 26 experts representing agencies that included those that provide data sharing resources for diseases prevalent in low and middle income countries.

What does this tell us?

Sharing health research data currently remains the exception rather than the norm. The review of research papers, including

completed clinical trials related to priority pathogens, found only 31% (98 out of 319 published papers, excluding case studies) provided access to all the data underlying the paper. While a few authors will provide the data on request, 65% of these papers give no information on how to find or access the data. And the review of clinical trial registries, for trials on interventions for priority pathogens, reported an even worse picture. Only two trials out of 58 provided any link in their registry entry to the background data[1,2].

Interviews with researchers revealed the reasons for a reluctance to share data included a lack of confidence in the utility of the data and therefore unwillingness to invest resources to prepare it to be shared; absence of academic-incentives for rapid dissemination that prevents subsequent publication (as opposed to the public health need) and a disconnect between those who are collecting the data and those who wish to use it quickly. A similar scepticism about how data might be used or misused, the potential harms to patients and the risks to the researcher sharing data that might reveal errors in their work, have been reported elsewhere[8,11].

Table 1 summarises the survey findings that identified which standards are used to share research data for neglected diseases and what those standards cover with respect to data curation, governance, security and longevity. Whilst there is clearly no universal or single standard to cover all the three areas and the elements under them, technical guidance is available across all the areas when those standards are combined. The standards created by the Clinical Data Interchange Standards Consortium (CDISC) were included as the United States Food and Drug Administration (FDA) has required the use CDISC standards in a clinical trial data submission since 2017. Hence these are widely used in industry and CDISC is fast becoming the *de facto* standard for data labelling and meta data.

What are the next steps: the role for funders in support of data sharing

Many health research funders have a generic policy requiring research data to be shared in a manner that maximises health and societal benefit. While some biomedical areas, like genomics, have forged ahead in maximizing data sharing, across health research more generally there is very low compliance with these

Table 1. Elements of interest addressed by data repository standards used in sharing data on the neglected diseases (adapted from Castillon *et al.*[9]).

AREAS	ELEMENTS	TRAC	ISO 16363	WHO	ICSU	H3Africa	CDISC
DATA CURATION	Metadata	+	+	-	-	+	+
	Discoverability	-	+	-	+	+	-
	Data Standardization	-	-	-	-	-	+
	Data Verification and Quality Assurance Procedures	+	-	+	+	+	-
SECURITY AND LONGEVITY	Encryption, Access Control & Other Security Measures	+	+	-	+	+	-
	Storage	+	-	-	+	+	-
	Data Backup	+	+	+	-	-	-
	Data Migration	+	+	-	-	-	-
	Sustainability of Funding (fee, free-to-access etc.)	+	+	+	+	-	-
	Data Preservation	-	+	-	+	-	-
	Succession Plan	+	+	+	-	-	-
GOVERNANCE	Legal Status	+	+	+	-	-	-
	Access and Terms of use	+	+	+	+	+	-
	Benefit sharing, Intellectual Property Issues	+	+	-	-	+	-
	Audit Procedures	+	+	+	-	-	-

+ As stated in publicly available information

- Information was not mentioned in the publicly available information.

TRAC (Trustworthy Repositories Audit & Certification), ISO 16363 (International Standards for Clinical Trial Registries -Space data and information transfer systems, Audit and certification of trustworthy digital repositories), WHO (World Health Organization), ICSU (International Council of Scientific Unions World Data System), H3Africa (The Human Heredity and Health in Africa Initiative), CDISC (Clinical Data Interchange Standards Consortium).

policies. In part this might reflect the limited guidance offered by the same funders in supporting their researchers to understand and undertake data sharing to implement and monitor these policies in practice.

It appears the main barrier to sharing is not technical but cultural, with researchers remaining sceptical about the benefits to them of sharing data. For researchers in low-resource setting data sharing can even be seen as a threat that their data will be exported and exploited by others with little benefit returning to them.

Therefore, research funders should take stock and revise data sharing policies to provide incentive structures for researchers. One clear first step would be to engage early with the researchers and related stakeholders to understand their concerns and work harder to define the benefit of sharing beyond a general sense that sharing data is in the public interest. The overall purpose of sharing the data needs to be clear and ideally developed with input from data suppliers, secondary data users, potential end-users and beneficiaries, and if possible with input from the participants that are the source of those data. Concerns regarding privacy versus the secondary use of the data need to be explored and mechanisms put in place to balance the public benefit against potential risks to privacy and confidentiality.

Secondly, there needs to be a direct benefit to sharing data that is directly relevant to those people that collect and curate the data. For example academics require citation of their work, including a data set. The generation of data and its subsequent citation for reuse needs to be integrated into research assessment – an idea captured in the Declaration on Research Assessment (see San Francisco Declaration on Research Assessment). So if the purpose of the data sharing mechanism is clear and all stakeholders buy into that purpose and if they feel their inputs will be recognised in research assessment together this will create a strong incentive to share. This was certainly our experience when working with Schistosomiasis researchers[8].

Thirdly, whilst there are a myriad of data standards to work with to meet the general principles of making data FAIR, more work needs to be done to realise the intent of making data sharing resources more equitable, ethical and efficient. As evident in the surveys summarized here good practice is starting to emerge so what is needed is better ways to share that practice. Funders need to work with the researchers and their networks to support the technical work required to develop standards that enable inter-operability.

For example one contributory role for funders would be to collect more systematically the data management plans that they have requested as part of funding grants and make them publicly accessible. In line with good practice these should be standardized where possible and ideally have clear, machine-readable metadata. An online resource that brings together the reference material and policies that are exemplars in each of the categories that cover governance, data curation, security and longevity would provide the basis for a framework to guide the future development of new sharing resources.

Finally, a checklist of the issues that need to be addressed when designing new or revising existing data sharing resources should be created. In addition to defining the purpose of data sharing this would highlight the technical, cultural and ethical issues that need to be considered and point to examples of emerging good practice that can be used to address them. The authors are working on this next stage and hope that with this type of planning and support in place the data sharing long desired by research funders will start to become the norm.

Data availability

No data are associated with this article.

Grant information

The author(s) declared that no grants were involved in supporting this work.

References

1. Pisani E, Ghataure A, Merson L: **Data sharing in public health emergencies: A study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats**. Wellcome Trust. 2018.
Publisher Full Text
2. Pisani E, Ghataure A, Merson L: **Supporting Data for: Data sharing in public health emergencies: A study of current policies, practices and infrastructure supporting the sharing of data to prevent and respond to epidemic and pandemic threats**. Harvard Dataverse, V1. 2017.
Publisher Full Text
3. Yozwiak NL, Schaffner SF, Sabeti PC: **Data sharing: Make outbreak research open access.** *Nature.* 2015; **518**(7540): 477–9.
PubMed Abstract | Publisher Full Text
4. Pisani E, Aaby P, Breugelmans JG, *et al.*: **Beyond open data: realising the health benefits of sharing data.** *BMJ.* 2016; **355**: i5295.
PubMed Abstract | Publisher Full Text
5. Olliaro PL: **Initiation and publication time-lags of treatment trials for Ebola virus disease.** *Lancet Infect Dis.* 2018; **18**(1): 28–29.
PubMed Abstract | Publisher Full Text
6. Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship.** *Sci Data.* [Internet]. 2016; **3**: 160018, [cited 2017 Mar 13].
PubMed Abstract | Publisher Full Text | Free Full Text
7. Hate K, Meherally S, Shah More N, *et al.*: **Sweat, Skepticism, and Uncharted Territory: A Qualitative Study of Opinions on Data Sharing Among Public Health Researchers and Research Participants in Mumbai, India.** *J Empir Res Hum Res Ethics.* 2015; **10**(3): 239–250.
PubMed Abstract | Publisher Full Text | Free Full Text
8. **Schistosomiasis Data Platform Stakeholder Meeting Report**. World Health Organization, Geneva. 2015.
Reference Source
9. Castillon G, Castilloux AM, Moride Y: **Development of Standards for Online Repositories**. Wellcome Trust, Version 3. 2017.
Reference Source

10. **Sharing health research data in low resource settings: Supporting necessary infrastructure and building on good practice.** Pisani E. Report prepared for Wellcome Trust and TDR the Special Programme for Research and Training in Tropical Diseases. 2018.
Publisher Full Text

11. Committee on Population; Division of Behavioral and Social Sciences and Education; The National Academies of Sciences, Engineering, and Medicine: **Sharing Research Data to Improve Public Health in Africa: A Workshop Summary**. Washington, DC: The National Academies Press, 2015.
PubMed Abstract | Publisher Full Text

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research