**Data sharing policies**

**TITLE**

Data sharing policies

**INTRODUCTION**

**Background**

The in-principle support for data sharing is strong. Research participants, biomedical researchers, healthcare consumers and health professionals are generally supportive of sharing of health research, clinical trial and health administrative, particularly if de-identified or anonymised.(1-8) The scientific community has gradually shifted from a cultural perception of data ownership to data custodianship and stewardship, and technological developments and statistical advances have greatly strengthened data sharing possibilities.(9)   
  
However, despite scientific and ethical imperatives, the in-practice commitment to sharing data is low.(10, 11) The practice of data sharing is dictated by data sharing policies. Over the past few decades, a large number of data sharing policies have been developed by a wide range of stakeholders. These data sharing policies are defined by a variety of principles and describe a variety of approaches to addressing issues.(12) However, they have been generally unsuccessful in closing the gap between the high level of in-principle support for data sharing and low level of in-practice willingness to actual do so.(13, 14)

There have been a few studies of data sharing policies. Blasimme et al (2018) conducted a network and qualitative content analysis of 230 data sharing policy documents by 97 organisations over two decades and found the data sharing policy landscape to be fragmented and stakeholder interests and expectations to be divergent.(12) They concluded that the research community should adopt and develop innovative tools, recalibrate and redistribute policy emphasis, and harmonise international data sharing policies.(12) Waithira et al. described the elements of a data sharing policy for institutions, departments or groups.(15) They included aims, data management for data sharing, models of sharing, data access criteria, consent models for participants, and budgeting and cost recovery.

Data sharing standards refer to agreed principles and recommended processes for the sharing of data between producers and custodians of data and other users of data, which accommodate variation in individual and group preferences (e.g., research and participant communities).(16). Data sharing standards differ from data sharing policies. Data sharing policies are specific to the aims, interests, expectations and context of stakeholder, sensitive to the regulatory requirements and ethical guidelines of the environment, and developed in consultation and engagement with internal and external groups. Data sharing standards, on the other hand, would support existing data sharing architecture (e.g., legal requirements and organisational governance), inform revisions of current data sharing policies and be incorporated into new data sharing policies.

**Objectives**

The aim of this paper is to review existing studies on data sharing policies and summarise the key findings at a stakeholder level. These key findings will include whether data sharing policies exist, whether they recommend or require data sharing, and what are the characteristics of existing policies. We have included studies of data sharing policies by various stakeholders in data sharing that sit at different stages in the life cycle of clinical trials. These are summarised in Figure 1.

Diagram

Description automatically generated **Figure 1.** Data sharing stakeholders by clinical trial stage

**HEALTH RESEARCH FUNDERS**  
Data sharing policies by health research funders which require data sharing have a large effect on facilitating data sharing and are associated with a high degree of compliance.(9) Health research funders which require data from health research they fund to be shared would minimise research waste and maximise research utility. They would minimise research waste by preventing duplication of data collection activities (especially those which are large, expensive, resource intensive, or nonreplicable), and maximise research utility by supporting timely secondary use of data (i.e., prospectively allocating sufficient resources to data sharing activities) and prospectively planned management (e.g. collection, curation and storage) of data.  
  
We identified three cross-sectional studies of data sharing policies of health research funders. A cross-sectional study of the data sharing policies of commercial and non-commercial funders found that only 38% of non-commercial funders had a data sharing policy (of which 60% encouraged data sharing and 40% mandated data sharing) and only 41% of commercial funders had a data sharing policy.(14) A cross-sectional study of the data sharing policies of clinical trial funders in France found that only 29% had a data sharing policy, of which 89% supported data sharing and 11% mandated data sharing.(17) Of clinical trials funders in France with a data sharing policy, 33% specified the type of data shared and the mode of sharing data but restricted the sharing of data to researchers.(17) Another cross-sectional study of life science researchers found that approximately a third of grant reviewers placed no weight on data sharing plans in their reviews.(9)  
  
**RESEARCH ETHICS COMMITTEES**Research ethics committees which recommend data from research they review be shared would strengthen the social licence and public acceptability of health research, and support the production of high quality datasets. Data sharing standards can support research ethics committees, who often consider whether proposals for data sharing meet ethical and regulatory standards.(18)   
  
We did not identify any studies of data sharing policies of research ethics committees.  
 **CLINICAL TRIAL REGISTRIES**Clinical trial registries which require data from clinical trials they register to be shared would expediate the validation of research (e.g., reproduction and replication through complete re-analysis of data) and synthesis of evidence.  
  
We did not identify any studies of data sharing policies of clinical trial registries.  
**PEER-REVIEWED SCIENTIFIC JOURNALS**Peer-reviewed scientific journals which require data from articles they publish to be shared would enhance the transparency of findings (i.e., openness of information irrespective of outcome) and confidence in results (e.g., accuracy and validity). The Principles of Transparency and Best Practice in Scholarly Publishing (<https://doaj.org/apply/transparency/>), by the Committee on Publication Ethics (COPE), Directory of Open Access Journals (DOAJ), Open Access Scholarly Publishers Association (OASPA) and the World Association of Medical Editors (WAME), requires member journals to have a policy on data sharing and reproducibility, and the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (<http://www.icmje.org/recommendations/>), by the International Committee of Medical Journal Editors (ICJME), requires clinical trials submitted to member journals to have a data sharing statement in the manuscript and a data sharing plan in the trial registration.(19) Publisher conformance with these guidelines and best practices is required for a journal to be indexed in MEDLINE or PubMed Central by the National Library of Medicine (<https://www.nlm.nih.gov/medline/medline_policies.html#pubpractices>).

Additionally, data publication, citation and altmetrics have been proposed as possible mechanisms to recognise data generation and incentivise data sharing.(20, 21) As these alternative scholarly crediting systems are implemented, data standards relating to data curation are required to protect the reliability and value of published data and prevent novel forms of research misconduct. This is particularly important given that there is only one evidence-based incentive (using open data badges) to promote data sharing.(22) However, the impact of data sharing policies by peer-reviewed scientific journals depends on author compliance and journal enforcement,(23) with data availability in less than half of RCTs in two journals (BMJ and PLOS Medicine) with strong data sharing policies. Suboptimal data availability despite strong data sharing policies is related to inability to contact corresponding author, inadequate resources by authors to prepare datasets, different data sharing practices, and infrequent formal or informal sanctions for non-compliance.(9, 24)

We identified four cross-sectional studies of data sharing policies of peer-reviewed scientific journals. A cross-sectional study of life, health and physical science journals found that 44% had no data sharing policy, 18% had a data sharing policy which encouraged data sharing and 38% had a data sharing policy which expected or mandated data sharing.(25) A cross-sectional study of biomedical journals found that 32% had no data sharing policy, 9% had a data sharing policy which mentioned data sharing, 23% had a data sharing policy which encouraged data sharing, 9% had a data sharing policy which required data sharing but not as a condition of publication, and 12% had a data sharing policy which required data sharing as a condition of publication.(25) Additionally, it found that most data sharing policies had no specific guidance on the practices of data sharing. A cross-sectional study of biology, clinical science, mathematics, physics and social sciences journals found that 44% had no data sharing policy, 5% had a data sharing policy which mentioned data sharing, 41% had a data sharing policy which encouraged data sharing, 8% had a data sharing policy which required data sharing but not as a condition of publication, and 3% had a data sharing policy which required data sharing as a condition of publication.(26) A cross-sectional study of ten high-impact surgical journals found that only one journal had a data sharing policy which required data sharing.(27)

Higher strength of data sharing policy is associated with higher impact factor.(25, 26, 28) Strength of data sharing policy is not associated with model of publication (open access or subscription).(28)  
  
**RESEARCH DATA REPOSITORIES**Research data repositories and other online infrastructure have a large effect on facilitating data sharing.(9) Research data repositories which recommend data they store be shared would motivate data sharing and improve data governance.   
  
We did not identify any studies of data sharing policies of research data repositories.

**OTHER**

Clinical trialist, research groups, academic departments and research institutions which require data they produce to be shared would improve the findability, accessibility, interoperability and reusability (FAIR principles) of their data, and the output of their research. They would support data producers by specifying agreed levels of data curation (e.g., collecting and organising data in a clear and consistent way) and anonymisation (e.g., reducing the risk of re-identification by eliminating direct identifiers and superfluous data and modifying indirect identifiers).(29) They would support data custodians by specifying agreed levels of data management (e.g., determining data access, use and exchange by validating secondary users research background and question, implementing data use agreements and approval processes, requiring statistical software programs to open datasets, or depositing data in research community accessible repositories).(29) They would also support data users with secondary analyses, study replication and individual participant data analysis. This could involve setting a minimum list of information for shared data to be understood, original analyses to be replicated, and raw data to be included in individual participant data analyses. Clinical trialists may also benefit from data sharing through additional scholarly citations (23) and research impact, or additional research collaborations and funding opportunities.  
  
However, in our retrospective cohort study of interventional trials, we found that only one in five committed to sharing data.(13) A cross-sectional study of trialists who were willing to share data found that approximately half had a data sharing plan, of which approximately half were written and half were discussed.(30)

**FUTURE DIRECTIONS**

The future directions of data sharing policies by themes are summarised in Table 1.

|  |  |
| --- | --- |
| **Table 1. Future directions of data sharing policies by themes\*** | |
| **Data sharing policy themes\*\*** | **Future directions** |
| Accountability | **Impact assessment guidance**  Researchers could more effectively conduct impact assessments for privacy or data protection (12)  **Participatory data governance**  Through participant-centred data governance mechanisms and models, researcher participants should be involved in decisions and control of data management, data access and data sharing, and included in data access committees and oversight bodies (12) |
| Attribution | **Data tagging systems**  Researchers should trace credit for data curation (12)  **Data citation incentives**  Research funders should recognise and reward data curators for their role in data sharing (12)  **Recognition**  Researchers could be required to recognise data producers through authorship, acknowledgement or citation |
| Capacity building |  |
| Common good |  |
| Data publication |  |
| Data quality and curation | **Reciprocity-based data-access models**  Through reciprocity-orientated data infrastructure, researchers who access shared data should contribute to its maintenance (e.g., conduct quality-related tasks), curation and development (e.g., contribute new data) (12) |
| Data access |  |
| Development |  |
| Group rights |  |
| Autonomy | **Digital consent technologies**  Through robust informed consent procedure, research participants could make meaningful, granular, case-by-case deliberations throughout the various uses of the data (12) |
| Integrity | **Collaboration**  Researchers should collaborate with data producers when the interpretation of the data requires the experience and knowledge of the data producers |
| Interoperability | **Interoperable data standards**  Research funders should reimburse or rewarding scientific institutions for sharing data with other research organisations (12) |
| Open access publication |  |
| Privacy | **Privacy-preserving technological solutions** (e.g., advanced cryptography techniques and distributed ledger technologies)  Through novel privacy preserving data sharing mechanisms and standards for minimum levels of consent management, data exchange and access-control policy enforcement, researchers could more effectively protect data security and information privacy, promote collaborative research through data sharing, and maintain transparency, traceability and immutability (12, 31-34) |
| Professionalism |  |
| Public engagement |  |
| Regulatory compliance | **Coherent regulatory environment**  Research community should create an integrated, coherent regulatory environment and system framework around the collection, use and distribution of data (12) |
| Risk benefit assessment |  |
| Solidarity |  |
| Sustainability |  |
| Timeliness |  |
| \*Proposals should be introduced through pilot initial and ad hoc data sharing policies and will likely require targeted public investment \*\*Identified by Blasimme et al (2018)(12) | |

**FUNDING**

There are no sources of financial support for this review.

**REFERENCES**

1. Hutchings E, Loomes M, Butow P, Boyle FM. A systematic literature review of attitudes towards secondary use and sharing of health administrative and clinical trial data: a focus on consent. Systematic Reviews. 2021;10(1).

2. Hutchings E, Loomes M, Butow P, Boyle FM. A systematic literature review of health consumer attitudes towards secondary use and sharing of health administrative and clinical trial data: A focus on privacy, trust, and transparency. Systematic Reviews. 2020;9(1).

3. Hutchings E, Loomes M, Butow P, Boyle FM. A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data. Systematic Reviews. 2020;9(1).

4. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. Journal of medical ethics. 2019;12.

5. Moon LA. Factors influencing health data sharing preferences of consumers: A critical review. Health Policy and Technology. 2017;6(2):169-87.

6. Garrison NA, Sathe NA, Antommaria AHM, Holm IA, Sanderson SC, Smith ME, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genetics in Medicine. 2016;18(7):663-71.

7. Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: A systematic literature review. Expert Review of Molecular Diagnostics. 2014;14(8):1053-65.

8. Eberlen JC, Nicaise E, Leveaux S, Mora YL, Klein O. Psychometrics Anonymous: Does a Transparent Data Sharing Policy Affect Data Collection? Psychol Belg. 2019;59(1):373-92.

9. Pham-Kanter G, Zinner DE, Campbell EG. Codifying collegiality: Recent developments in data sharing policy in the life sciences. PLoS ONE. 2014;9(9).

10. Ramagopalan SV, McDonald L, Schultze A, Simpson A, Graham S, Wasiak R. A review of data sharing statements in observational studies published in the BMJ: A cross-sectional study. F1000Research. 2017;6 (no pagination).

11. Gorman DM. Availability of Research Data in High-Impact Addiction Journals with Data Sharing Policies. Science and engineering ethics. 2020;26(3):1625-32.

12. Blasimme A, Fadda M, Schneider M, Vayena E. Data sharing for precision medicine: Policy lessons and future directions. Health Affairs. 2018;37(5):702-9.

13. Tan AC, Askie LM, Hunter KE, Barba A, Simes RJ, Seidler AL. Data sharing—trialists' plans at registration, attitudes, barriers and facilitators: A cohort study and cross-sectional survey. Research Synthesis Methods. 2021;12(5):641-57.

14. Gaba JF, Siebert M, Dupuy A, Moher D, Naudet F. Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. PLoS ONE. 2020;15(8):e0237464.

15. Waithira N, Mutinda B, Cheah PY. Data management and sharing policy: The first step towards promoting data sharing. BMC Medicine. 2019;17(1).

16. Robinson JO, Slashinski MJ, Chiao E, McGuire AL. It depends whose data are being shared: considerations for genomic data sharing policies. J Law Biosci. 2015;2(3):697-704.

17. Rollando P, Parc C, Naudet F, Gaba JF. Data sharing policies of clinical trials funders in France. Therapie. 2020.

18. Lemke AA, Smith ME, Wolf WA, Trinidad SB, Consortium G. Broad data sharing in genetic research: views of institutional review board professionals. Irb. 2011;33(3):1-5.

19. International Committee ofMedical Journal Editors’. Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals 2019. p. 14.

20. Devriendt T, Shabani M, Borry P. Data Sharing in Biomedical Sciences: A Systematic Review of Incentives. Biopreservation and biobanking. 2021;11.

21. Poline JB. From data sharing to data publishing [version 2; peer review: 2 approved, 1 approved with reservations]. MNI Open Res. 2019;2.

22. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. Res. 2017;2:4.

23. Christensen G, Dafoe A, Miguel E, Moore DA, Rose AK. A study of the impact of data sharing on article citations using journal policies as a natural experiment. PLoS ONE. 2019;14(12).

24. Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in the BMJ and PLOS Medicine. BMJ (Online). 2018;360 (no pagination).

25. Kim J, Kim S, Cho HM, Chang JH, Kim SY. Data sharing policies of journals in life, health, and physical sciences indexed in Journal Citation Reports. PeerJ. 2020;8 (no pagination).

26. Resnik DB, Morales M, Landrum R, Shi M, Minnier J, Vasilevsky NA, et al. Effect of impact factor and discipline on journal data sharing policies. Accountability in research. 2019;26(3):139-56.

27. Chapman SJ, Shelton B, Mahmood H, Fitzgerald JEF, Harrison E, Bhangu A. Promoting transparency in clinical research: Systematic review of disclosure and data-sharing policies in surgical journals. International Journal of Surgery. 2014;12:S52.

28. Vasilevsky NA, Minnier J, Haendel MA, Champieux RE. Reproducible and reusable research: Are journal data sharing policies meeting the mark? PeerJ. 2017;2017(4).

29. Rodriguez A, Tuck C, Dozier MF, Eguiagaray IM, Eldridge S, Lewis SC, et al. Current recommendations/practices for anonymising data from clinical trials in order to make it available for sharing: A scoping review. Trials Conference: 5th International Clinical Trials Methodology Conference, ICTMC. 2019;20(Supplement 1).

30. Tannenbaum S, Ross JS, Krumholz HM, Desai NR, Ritchie JD, Lehman R, et al. Early experiences with journal data sharing policies: A survey of published clinical trial investigators. Annals of Internal Medicine. 2018;169(8):586-8.

31. Rezaeibagha F, Mu Y. Distributed clinical data sharing via dynamic access-control policy transformation. International Journal of Medical Informatics. 2016;89:25-31.

32. Dubovitskaya A, Novotny P, Xu Z, Wang F. Applications of Blockchain Technology for Data-Sharing in Oncology: Results from a Systematic Literature Review. Oncology (Switzerland). 2020;98(6):403-11.

33. Sarwate AD, Plis SM, Turner JA, Arbabshirani MR, Calhoun VD. Sharing privacy-sensitive access to neuroimaging and genetics data: A review and preliminary validation. Frontiers in Neuroinformatics. 2014;8(APR).

34. Jiang X, Sarwate AD, Ohno-Machado L. Privacy technology to support data sharing for comparative effectiveness research: A systematic review. Medical Care. 2013;51(8 SUPPL.3):S58-S65.

35. Bredenoord AL, Mostert M, Isasi R, Knoppers BM. Data sharing in stem cell translational science: Policy statement by the International Stem Cell Forum Ethics Working Party. Regenerative Medicine. 2015;10(7):857-61.

**NON-REFERENCES**

*Background*1. YODA Project (<https://yoda.yale.edu/relevant-literature>)

2. Data Champions Project (<https://www.data.cam.ac.uk/intro-data-champions>)   
3. Research data alliance (<https://www.rd-alliance.org>)   
- Hrynaszkiewicz et al (2018) Research Data Alliance (RDA) Data Policy Standardisation and Implementation Working Group Research Data Alliance. Data policy standardisation and implementation [Internet]. The Alliance; 2017 [cited 14 Dec 2017]. <https://www.rd-alliance.org/groups/data-policystandardisation-and-implementation>.  
- Reproducible Health Data Services WG (<https://www.rd-alliance.org/node/61938/case-statement>)

- Raising FAIRness in health data and health research performing organisations (HRPOs) WG (<https://www.rd-alliance.org/node/69831/case-statement>)   
- Research Data Alliance (RDA) Data Policy Standardisation and Implementation Working Group  
- RDA Privacy Implications of Research Data Sets IG (<https://www.rd-alliance.org/node/50796/charter>)   
- Health Data Interest Group (<https://www.rd-alliance.org/node/50708/charter>)   
4. FAIRsharing (<https://fairsharing.org>)   
5. Force 11

- FAIR Data Principles (<https://www.force11.org/group/fairgroup/fairprinciples>)

- Research Data Publishing Ethics (<https://www.force11.org/group/research-data-publishing-ethics>)

6. CDISC standards (<https://www.cdisc.org/standards/foundational/cdash>)  
7. Sherpa Juliet (<https://v2.sherpa.ac.uk/juliet/about.html>)

8. Coalition for Publishing Data in the Earth and Space Sciences (<https://copdess.org/>)

9. DISTILL project shared by Lene on OneDrive (<https://unisyd-my.sharepoint.com/personal/lene_seidler_sydney_edu_au/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Flene%5Fseidler%5Fsydney%5Fedu%5Fau%2FDocuments%2FData%20sharing%20project&ct=1635925368941&or=OWA%2DNT&cid=6c0a8b5c%2De84c%2D6312%2D1d57%2D57cd99541c51>)

10. The REPRISE project  
<https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-021-01670-0>   
11. Rates and predictors of data and code sharing in the medical and health sciences: Protocol for a systematic review and individual participant data meta-analysis  
<https://f1000research.com/articles/10-491/v2> *Health researcher funders*  
1. NHMRC Open Access Policy  
[https://www.nhmrc.gov.au/about-us/resources/open-access-policy?utm\_medium=email&utm\_campaign=Tracker%20-%201%20November%202021&utm\_content=Tracker%20-%201%20November%202021+CID\_df6913d5612e708479ca77004b5081b9&utm\_source=Mailbuild&utm\_term=report#](https://www.nhmrc.gov.au/about-us/resources/open-access-policy?utm_medium=email&utm_campaign=Tracker%20-%201%20November%202021&utm_content=Tracker%20-%201%20November%202021+CID_df6913d5612e708479ca77004b5081b9&utm_source=Mailbuild&utm_term=report)   
2. NIH Data Sharing policy  
3. Wellcome Trust data sharing policy  
4. Gates Foundation data sharing policy  
  
*Other*  
In the fields of genomics, the Global Alliance for Genomics and Health has adopted the Framework for Responsible Sharing of Genomic and Health-Related Data.(35)