Publishing descriptions of non-public clinical datasets: guidance for researchers, repositories, editors and funding organisations

Iain Hrynaszkiewicz[1], Varsha Khodiyar[2], Andrew L Hufton[2] and Susanna-Assunta Sansone[2,3]

[1]Nature Publishing Group & Palgrave Macmillan, The Macmillan Campus, Trematon Walk Wharfdale Road, London N1 9FN, United Kingdom

[2]*Scientific Data*, The Macmillan Campus, Trematon Walk, Wharfdale Road, London N1 9FN, United Kingdom

[3]Oxford e-Research Centre, University of Oxford, Oxford OX1 3QG, United Kingdom

Abstract

Sharing of clinical research data usually happens between individuals or research groups rather than via public repositories due to the need to protect research participant privacy. This approach to data sharing makes it difficult to connect journal articles with their underlying datasets and is often insufficient for ensuring access to data in the long term. The Yale Open Data Access (YODA) and Clinical Study Data Request (CSDR) projects have increased accessibility to clinical datasets for secondary uses while protecting patient privacy and the legitimacy of secondary analyses but these resources are generally disconnected from journal articles – where researchers typically search for reliable information to inform future research. New types of journal and journal article dedicated to publishing data articles have emerged in recent years and journals are developing stronger links with data repositories. There is a need for increased collaboration between journals, data repositories, researchers and their sponsors, and "data on request" services such as YODA and CSDR to increase the visibility and reliability of clinical research. We propose changes to the format and peer-review process for journal articles to more robustly link them to data that are available on request. We also propose additional features for data repositories to better accommodate non-public clinical datasets, including Data Use Agreements (DUAs).

Introduction

Open access to research data that can be understood and reused by others is a means to further scientific progress and publish more reliable and reproducible research[1,2]. However, clinical research data often include information that could potentially identify individuals, meaning datasets must be anonymised prior to being shared beyond the study for which the data were originally collected. Processes for anonymising human data must be rigorously applied to maintain individual privacy, scientific value and data integrity. Although guidelines and processes for anonymisation of clinical data exist,[3,4] publication of freely available clinical datasets (such as[5]) remains uncommon. As open access to clinical datasets is often unfeasible, a more felicitous and pragmatic approach may be needed.

Some clinical datasets may be made available on request from authors of research articles or through the recent emergence of dedicated data request systems. However, as large amounts of clinical research can go unpublished[6,7], and clinical trials unregistered[8], the discoverability of clinical datasets is suboptimal. In this paper we use the term "non-public clinical datasets" to refer to datasets that are not openly accessible, but are available on request.

In consultation with interested parties representing pharmaceutical companies, research funders, researchers and data repositories, the editors and publishers of the journal *Scientific Data* propose guidelines for linking peer-reviewed journal articles to non-public clinical datasets.

Data repositories are essential for enabling reliable access to data underlying research. Required features of data repositories for non-public datasets, to enable linkage of these datasets to journal articles, are also discussed. We also propose criteria by which repositories for non-public clinical datasets could be assessed, to determine if a repository can host non-public data permanently and can establish robust links with the peer-reviewed literature.

Summary of recommendations

Clinical researchers and their sponsors

- *Be prepared to share data with editors, peer reviewers and other researchers in accordance with journal policies*
- *Apply the shortest possible embargoes on data*

Repositories

- *Develop mechanisms to host clinical research data, including:*
    - *Provide stable identifiers for metadata records about non-public dataset(s)*
    - *Implement Data Use Agreements (DUAs)*
    - *Implement a transparent system for requesting access to data and reviewing requests to access data*
    - *Allow access to data in a timely manner and include a proportionate review of the scientific rationale, without introducing unnecessary barriers*

Clinical journal editors and publishers

- *Check compliance with journal data sharing policies for every submission*
- *For manuscripts based on secondary access to trial data (e.g. data the original trial sponsor has made available for further research), check the research is consistent with the DUA and purpose for which data access was granted*
- *Facilitate peer review of clinical datasets more systematically*
- *Build relationships with repositories for non-public clinical datasets*
- *Introduce data sharing statements and transparency statements in published articles*

Data journal editors and publishers

- *Develop data article (Data Descriptor) formats to permanently link articles to descriptions of non-public clinical datasets*

All sponsors and funders of trials

- *Build partnerships with and between data sharing initiatives, trusted repositories, and peer-reviewed journals committed to data sharing*
- *Apply the shortest possible embargoes on data and ensure that data access is subject to a proportionate review (e.g. of the scientific rationale and qualifications of the research team), without introducing unnecessary barriers*

How do researchers currently access non-public clinical data?

Data sharing between researchers has traditionally occurred through direct contact between individuals and research groups[9]. Many journals have policies that require authors to share data that support their results with other researchers on request, but enforcement of such policies varies between journals (for a summary of journal policy types and approaches see[10]). Journal policies that only require data to be "available on request" without also mandating data availability statements, have been found to be less effective for ensuring data access for future researchers[11–14]. However, even journals with strong and enforced policies on data access, (such as at *Nature Medicine*, the *BMJ* and *PLOS Medicine)*, data about identifiable human subjects will usually not be in the public domain, due to the need to protect research participants' privacy.

Alongside changes in journal policies, initiatives from other stakeholders (regulatory agencies, the pharmaceutical industry and research groups) in clinical research have begun to facilitate greater access to non-public clinical datasets.

In January 2015 the European Medicines Agency (EMA) put into place a policy on publishing clinical reports (documents, rather than data), submitted as part of marketing-authorisation applications for human medicines. These documents will be made available via a website that requires registration and agreement to terms of use. In 2015 the agency is continuing to develop its policy, which will in the future provide access to individual patient data (IPD)[15].

When surveyed about their data sharing attitudes and behaviours, clinical researchers express concerns about inappropriate secondary analysis of their data, and they cite concerns about patient privacy[16]. The Yale Open Data Access (YODA; http://yoda.yale.edu/) project and Clinical Study Data Request (CSDR; http://clinicalstudydatarequest.com) have, however, since 2012, provided restricted access to non-public clinical datasets while addressing these two concerns. As of March 2015, CSDR listed more than 2100 clinical studies from 12 pharmaceutical companies, for which access to data could be requested. Researchers are also able to enquire about the availability of other non-listed studies. The project has been described as a success by its independent data access review panel[17]. As of March 2015 GlaxoSmithKline, one of the companies using CSDR, has received 99 requests (approximately four requests per month) to access data from the 1200 trials it has listed, and the YODA Project has compiled data from 112 studies, representing two commercial data providers, and has received more than 10 requests to access data.

Connecting data available on request with journals and repositories

With increasing numbers of open access repositories for research data for many scientific disciplines (http://www.nature.com/sdata/data-policies/repositories) that provide persistent, citable links to datasets and metadata records (landing pages) about datasets, it is relatively easy to link journal articles to publicly accessible datasets. In the area of clinical trials, steps to provide more routine access to data are relatively new and data are not yet widely available from across all parts of the research community. In addition, links between the peer-reviewed literature and non-public clinical datasets are far less robust than links between literature and public datasets. There is a need to integrate initiatives such as YODA and CSDR with journal policies designed to integrate data and literature to support more reproducible research. This could be achieved through developing the data access policies of journals and developing the relationships of journals with data repositories that hold non-public datasets. It could also be achieved by a new type of journal article for describing non-public clinical datasets. These new articles would be an adaptation of "data papers" – or, in the case of *Scientific Data*, Data Descriptors – that are published in data journals. Data journals are a relatively new type of journal focused on data publication[18]. Below we describe the benefits of this approach for researchers, opposed to simply posting information about non-public datasets on a website or posting summary results of clinical trials in trial registration databases.

Discoverability

YODA and CSDR provide public information about clinical studies for which data are available on request. Some of these studies may not be described in journal articles and for studies that *are* published it may not be clear in the published article that the underlying data are available. Moreover, when planning a new research project or systematic review, medical researchers primarily rely on bibliographic databases, journals and clinical trial registries to find information rather than these, relatively new, websites. A possible solution is data journals and data articles, which provide a way to publish, in a peer-reviewed journal, discoverable – via bibliographic and other databases – detailed descriptions of research datasets. This is also a means to give credit to data generators through publications.[9].

Scientific Data (http://www.nature.com/sdata/) (Nature Publishing Group) is one example of a data journal (see[18] for others) – an open access journal for descriptions of scientifically valuable datasets, where articles (termed Data Descriptors) are linked to their corresponding publicly-available datasets. With non-public datasets an important difference is that articles need to link to a persistent and citable online metadata record about the dataset, a "stub" record or landing page, for a clinical dataset that is available on request. Data Descriptors might be written by those who created the non-public clinical dataset as a means to get more credit, and provide greater incentive for providing data access. This might also be an option for researchers granted secondary access to data whose study of another researchers' data did not fit a traditional research article format.

As well as increasing discoverability of content through indexing in bibliographic databases, such as PubMed/MEDLINE and Google Scholar, open access journals, in particular, can help ensure content is highly visible through exposure to standard internet search engines such as Google.

Quality and peer review

Bibliographic databases, particularly the more selective databases such as PubMed/MEDLINE, Web of Science and Scopus, give some assurances of quality and reliability of the included information to their users through journal evaluation and selection procedures.

Peer review is central to publishing research in journals. Publication of Data Descriptors at *Scientific Data* (and some other data journals), involves formal peer review by independently-selected experts, of both the article describing the dataset(s) and the dataset itself. The peer-review process at *Scientific Data* focuses on (re)usability and data integrity, rather than on the perceived importance or impact of the data (http://www.nature.com/sdata/for-referees). Peer review of underlying data systematically is not, however, routine in traditional research journals.

Data curation

Scientific Data's publication process includes data curation by a dedicated Editor, in addition to peer reviewer checks. This process includes the creation of standardised, machine readable metadata for every Data Descriptor. This is intended to facilitate data discoverability and reuse by using controlled vocabulary terms to indicate sample and subject provenance and outline the experimental workflow (http://www.nature.com/sdata/about/faq#q14). Datasets in curated, readily consumable formats should increase confidence in the delivered data formats, adding further value to the offerings of data journals.

Permanence

A major role of journals is to ensure the permanence and integrity of the scientific record, for example by maintaining persistent links between articles and datasets, and placing copies of content in redundant archives. Web link decay or "link rot" is well documented, and even in the peer-reviewed literature, an estimated 20% of articles published in 2012 already suffer from broken web links when regular websites or URLs are cited[19]. This deterioration of referential integrity across corresponding data sources presents obstacles to replicating or re-analysing results underlying the scientific record as the outputs of research studies (primarily datasets) cannot be easily located. Publishers' use of persistent identifiers for journal articles, Digital Object Identifiers (DOIs), helps ensure readers and future researchers can access content as publishers commit to keeping article links up-to-date using the DOI system. DOIs are increasingly created for datasets and metadata records, via data repositories.

Links with data repositories

While the YODA Project and CSDR have succeeded in increasing access to clinical research data, their websites, and documents hosted therein are not citable and linkable in the same way as journal articles, and other research objects assigned DOIs. Furthermore, researchers and projects funded by organisations such as Cancer Research UK, Medical Research Council Clinical Trials Unit (MRC CTU)[20] and the Wellcome Trust may have well-managed non-public datasets available via local hosting, but these archives are unlikely to meet the repository assessment and linking criteria of journals and publishers. *Scientific Data*, for example, works with trusted repositories to publish its Data Descriptors and supports data archiving policies and activities across the *Nature* journals[2]. *Scientific Data* has established criteria for assessing public research data repositories (see below) and has so far approved more than 70 repositories (http://www.nature.com/sdata/data-policies/repositories). Other publishers and journals also list suggested or recommended repositories for authors including BMJ Open, PLOS, BioMed Central.

Additional considerations for journals and data repositories

Data repositories for non-public clinical datasets need to provide additional services to those provided by public data repositories. These criteria, outlined below, apply equally for repositories wishing to link to Data Descriptors or to traditional clinical research articles.

Data use agreements

An essential component of secondary use of non-public datasets is a data use agreement (DUA) between the data generator or repository and the secondary researcher(s). The purposes of DUAs include reducing risks to participants and other parties involved in the study and to ensure the scientific value of secondary analyses. The Institute of Medicine's (IOM) Committee on Strategies for Responsible Sharing of Clinical Trial Data 2015 report[21] (Chapter 5, p118) lists as characteristics of DUAs:

"*Common provisions of DUAs to reduce risks:*

- *Prohibit re identification attempts*
- *Prohibit further sharing*
- *Prohibit use to support a competing license application*
- *Acknowledging original trialists*
- *Assignment of IP*

Provisions to enhance scientific value:

- *Publish analysis in peer reviewed journal and make statistical analysis plan available*
- *Send copies of manuscripts to original investigators/sponsors with no right of revision*

- *Restrictions on reuse not specified in original application"*

DUAs are a part of the YODA Project and CSDR processes (for example, https://www.clinicalstudydatarequest.com/Documents/DATA-SHARING-AGREEMENT.pdf). Another example is the National Cancer Research Institute (NCRI) which has prepared a template for biosample access policy development that enables biobanks to share materials with other researchers (http://www.ncri.org.uk/wp-content/uploads/2013/09/Initiatives-Biobanking-2-Access-template.pdf).

While a journal or data journal cannot require a particular wording of DUAs, it is our view that data must be available in a manner that allows qualified researchers to use the data in appropriate contexts, and permits critical reanalyses, including those which may conflict with the interests of the sponsors of the original study. We recommend that an approved repository for clinical data should include assessing DUAs to ensure there is an appropriate balance between restriction and reuse. Mandatory requirements for co-authorship for data creators should be avoided – to help ensure independence of secondary analyses.

Controlled access and governance

General research data repositories such as figshare, Dryad and Dataverse work with journals and publishers to link articles to supporting datasets. Specialist data repositories are usually the best place for the long term storage and curation of research data, with general data repositories providing a home for orphan datasets[2]. However, few repositories have processes for managing and approving requests to access non-public clinical datasets. These processes can include independent advisory committees, and are provided by CSDR and the YODA Project. Moreover, few repositories provide statistical analysis software (SAS/R/STATA) environments for reanalysis of data that is provided by CSDR and the YODA Project. Other benefits of the CSDR and the YODA Project approach include autonomy and independence of study sponsor approval and transparency of process – with public listing of requests to access data and the outcome of these requests[17] (https://www.clinicalstudydatarequest.com/Metrics.aspx). For CSDR, the Wellcome Trust, in April 2015, took responsibility for managing the review of research proposals and the operation of the Independent Review Panel. The role of a trusted intermediary for repositories for clinical data is also highlighted by the IOM's report. This suggests potential for greater collaboration between initiatives such as CSDR and the YODA Project, repositories, and journals.

Landing pages

Repository metadata records (landing pages) for linking journal articles to non-public clinical data will share many characteristics with repository records for public datasets. Good practice for data citation already recommends that links to datasets should resolve to landing pages rather than the raw data files, and standards are emerging for the information that is essential and desirable to be included in landing pages[22]. These standards consist of a dataset description comprising a dataset identifier, title and brief description, creator, Publisher, and publication year. Landing pages should also include persistence/permanence information and licensing information for the data. We recommend that landing pages for non-public clinical datasets also include information detailing the access controls pertinent to the data.

Additional repository assessment criteria

Taking the above issues into consideration, we propose additional criteria by which journals and publishers can assess repositories for hosting of non-public clinical datasets:

Trusted repositories for non-public datasets should:

- *Provides stable identifiers for metadata records ("landing pages") about non-public dataset(s)*
- *Allows access to data with the minimum of restrictions needed to ensure protection of privacy and appropriateness of secondary analyses, codified in Data Use Agreements (DUAs)*
- *Be independent of the study sponsors and principal investigators [and their institution(s)]*
- *Have a transparent and persistent system for requesting access to data and reviewing requests to access data*
- *Allow access to data in a timely manner*
- *Ensure long-term preservation of data in their non-public form*

These criteria are in addition to the current repository selection criteria (http://www.nature.com/sdata/about/faq#q21) of *Scientific Data* which require trusted repositories:

- *Be broadly supported and recognized within their scientific community*
- *Ensure long-term persistence and preservation of datasets in their published form*
- *Provide expert curation*
- *Implement relevant, community-endorsed reporting requirements*
- *Provide for confidential review of submitted datasets*
- *Provide stable identifiers for submitted datasets*
- *Allow public access to data without unnecessary restrictions*

Repositories that could or may meet the current and additional requirements

Specialist trial data systems such as CSDR (and YODA) have indicated a willingness to listen to feedback from researchers and may evolve as data sharing progresses. As well as the potential for these repositories to develop to meet certain criteria, there are number of other general repositories that could or may meet the above criteria:

- UK Data Archive, which manages access to "safeguarded" (http://ukdataservice.ac.uk/get-data/data-access-policy/safeguarded-data.aspx) and "controlled" (http://ukdataservice.ac.uk/get-data/data-access-policy/controlled-data.aspx) datasets
- Inter-university Consortium for Political and Social Research (ICPSR), which provides a virtual data enclave system and permanent metadata records (http://icpsr.blogspot.co.uk/2014/05/icpsrs-virtual-data-enclave-prepared-to.html)
- Dataverse, which is developing privacy tools for sharing research data (http://privacytools.seas.harvard.edu/; http://datascience.iq.harvard.edu/files/datascience/files/opendata-datatags-mercecrosas.pdf)
- Figshare (with software and governance development; disclosure: figshare were represented in our working group)

Additional manuscript sections required

The format of published articles – Data Descriptors and research articles – needs to be developed to accommodate links to non-public datasets.

Articles should include information about why the data are not publicly available (i.e. because they contain personally identifiable information) and describe the restrictions on accessing the dataset. They should also state if the data are subject to a DUA and where the DUA can be found – ideally, this should be hosted permanently with the landing page of the non-public dataset. In *Scientific Data*, the Usage Notes section would be appropriate for this information. Persistent links to landing pages should also be cited, in the same way public datasets are cited. Other journals, such as *PLOS ONE*, *Palgrave Communications*, *GigaScience* and *Royal Society Open Science* are now routinely including dedicated article sections to describe and link to datasets supporting published articles – these could be adapted to meet these requirements.

Where Data Descriptors, and other articles, link to non-public clinical datasets we recommend authors include in their articles a transparency declaration guaranteeing that their description of the dataset is an honest and accurate account. Transparency statements for regular journal articles, for other aspects of research integrity, have been implemented by the BMJ[23].

See Figure 1 for an overview of the standard editorial workflow of *Scientific Data*, and Figure 2 for the proposed modified workflow to accommodate Data Descriptors of non-public clinical datasets.

Research participant consent

Part of a journal's role is to enforce relevant ethical – and legal – expectations regarding consent. An important consideration is that participants gave appropriate consent for data to be made available to secondary researchers in the future. Where informed consent for data sharing was obtained, consent should include articles designed to describe or support the release of datasets.

What data should be available to secondary researchers?

Different types of experiments produce different types of information in a variety of formats, leading to different minimum requirements for secondary researchers seeking to replicate or understand results. In general, reproducible medical research requires access to data, code, and study protocols[24]. CSDR has, furthermore, defined data and document types for the studies it lists, although all items are not always available for each study:

i. Raw dataset
ii. Blank case report forms
iii. Annotations of blank case report forms
iv. Dataset specifications
v. Protocol (all versions)
vi. Analysis-ready dataset
vii. Reporting and analysis plan
viii. Clinical study report

While the scope of these guidelines potentially goes beyond clinical trials – including molecular data types – this list defined by CSDR is a reasonable guide. The IOM has also described the clinical research data types that are needed for reanalysis, which differs slightly from the CSDR list (Chapter 5, p112)[21]. Providing all these items might not be feasible in all cases however and, in the case of *Scientific Data*, the Editors and peer reviewers would focus on whether sufficient information has been provided to enable datasets to be reused, and if the data are scientifically valuable. Standards for reusability for historic, non-public clinical datasets might need to be less stringent if data are only available in file formats that might not be optimised for reuse.

Who should have secondary access to data?

The consensus of regulators, industry sponsors and funders of clinical trials is for data access to be granted to suitably qualified researchers with a legitimate reanalysis proposal. A requirement of CSDR's procedures is that a researcher with a degree in statistics or a related discipline should be part of the research team. The Independent Review Panel for CSDR had, as of February 2015, approved 71 requests and rejected or advised to resubmit 4 requests. The YODA Project's approval committee assesses "basic information about the Principal Investigator, Key Personnel, and the project Research Proposal, including a scientific abstract and research methods" when reviewing data access requests[25].

Journal polices generally require data supporting submitted works must be accessible to peer reviewers and editors, and study sponsors should already be used to providing access to data supporting manuscripts submitted to major medical journals. These repository criteria and guidelines could make these processes more efficient if applied to clinical research journals. To publish Data Descriptors in *Scientific Data*, peer reviewers and editors must be given controlled access to supporting data for every article. The majority of journals operate single or double blind peer review, which means some reviewers or journals might require their anonymity to be maintained (public data repositories often support anonymous peer review, although there is increasing adoption of open peer review).

When to provide secondary access to data

The IOM has recommended embargoes of up to 18 months from study completion before clinical trialists are required to share data, although this has been criticised for being too long[26]. In a major epidemic, a long embargo on data access and reuse could be to the detriment of fighting disease. As of 1st January 2015 the Bill and Melinda Gates Foundation, a funding organisation, introduced a policy of requiring immediate open access to data underlying published research, although this does not specifically relate to clinical trial datasets (http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy). In general, a reasonable period for analysis – a right of first use – is acknowledged in most research communities. Any embargo on non-public clinical dataset(s) described in and linked to a journal article would have to have expired to comply with the recommendations in this article. In general we advocate no, or short, embargoes on data release wherever feasible.

Next steps

In consultation with a Working Group, convened in December 2014, (see Acknowledgements) *Scientific Data* is developing its editorial and peer-review processes and relationships with repositories to support publication of Data Descriptors for non-public clinical datasets. Other journals – data journals and traditional journals – may wish to consider these repository, linking and editorial policy proposals. Some members of our working group are also helping to identify interested research teams and relevant datasets that could be part of a publication pilot. Indeed, we need real data with which to develop more robust links between non-public datasets and journal articles. We strongly encourage others to contact the editors (scientificdata@nature.com), to discuss proposals.

Acknowledgements

Thanks to the Working Group members, who attended a meeting in December in 2014 which identified many of the issues highlighted in this paper and who provided comments on earlier drafts of the paper:

Robert Frost, Policy Director, Medical Policy, GSK
Richard Lehman, on behalf of the Yale Open Data Access Project
Fiona Reddington, Head of Clinical and Population Research Funding, Cancer Research UK
Will Greenacre, Policy Officer, Wellcome Trust
Matthew Sydes, Senior Scientist and Statistician, MRC Clinical Trials Unit

John Gonzalez, Publications Director, AstraZeneca
Jay Bergeron, Director, Translational and Bioinformatics, Pfizer
Jessica Ritchie/Joseph Ross, Yale Open Data Access Project
Mark Hahnel, CEO and Founder, figshare

Competing interests

IH, VK and ALH are employees of Nature Publishing Group, which publishes *Scientific Data*. SS is Honorary Academic Editor of *Scientific Data*.

References

1. *Science as an open enterprise*. (Royal Society, 2012). at <https://royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe.pdf>

2. Data-access practices strengthened. *Nature* **515,** 312–312 (2014).

3. Hrynaszkiewicz, I., Norton, M. L., Vickers, A. J. & Altman, D. G. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ* **340,** c181–c181 (2010).

4. El Emam, K., Rodgers, S. & Malin, B. Anonymising and sharing individual patient data. *BMJ* **350,** h1139–h1139 (2015).

5. Sandercock, P. A. G., Niewada, M. & Członkowska, A. The International Stroke Trial database. *Trials* **12,** 101 (2011).

6. Jones, C. W. *et al.* Non-publication of large randomized clinical trials: cross sectional analysis. *BMJ* **347,** f6104–f6104 (2013).

7. McGauran, N. *et al.* Reporting bias in medical research - a narrative review. *Trials* **11,** 37 (2010).

8. Van de Wetering, F. T., Scholten, R. J. P. M., Haring, T., Clarke, M. & Hooft, L. Trial registration numbers are underreported in biomedical publications. *PLoS One* **7,** e49599 (2012).

9. Kratz, J. E. & Strasser, C. Researcher Perspectives on Publication and Peer Review of Data. *PLoS One* **10,** e0117619 (2015).

10. Hrynaszkiewicz, I., Li, P. & Edmunds, S. C. in *Implementing Reproducible Research* (eds. Stodden, V., Leisch, F. & Peng, R. D.) (CRC Press, 2014). at <http://books.google.co.uk/books?hl=en&lr=&id=JcmSAwAAQBAJ&oi=fnd&pg=PA383&ots=ylWdxQyJLE&sig=-bwQySeGBXkNVxg2cF-uN0pcE88#v=onepage&q&f=false>

11. Savage, C. J. & Vickers, A. J. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* **4,** e7078 (2009).

12. Vines, T. H. *et al.* Mandated data archiving greatly improves access to research data. *FASEB J.* fj.12–218164– (2013). doi:10.1096/fj.12-218164

13. Jaspers, G. J. & Degraeuwe, P. L. J. A failed attempt to conduct an individual patient data meta-analysis. *Syst. Rev.* **3,** 97 (2014).

14. Wicherts, J. M., Borsboom, D., Kats, J. & Molenaar, D. The poor availability of psychological research data for reanalysis. *Am. Psychol.* **61,** 726–728 (2006).

15. Koenig, F. *et al.* Sharing clinical trial data on patient level: opportunities and challenges. *Biom. J.* **57,** 8–26 (2015).

16. Rathi, V. *et al.* Clinical trial data sharing among trialists: a cross-sectional survey. *BMJ* **345,** e7570 (2012).

17. Data Sharing, Year 1 — Access to Data from Industry-Sponsored Clinical Trials — NEJM. at <http://www.nejm.org/doi/full/10.1056/NEJMp1411794>

18. Kratz, J. & Strasser, C. Data publication consensus and controversies. *F1000Research* **3,** 94 (2014).

19. Klein, M. *et al.* Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot. *PLoS One* **9,** e115253 (2014).

20. Sydes, M. R. *et al.* Sharing data from clinical trials: the rationale for a controlled access approach. *Trials* **16,** 104 (2015).

21. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk - Institute of Medicine. at <http://www.iom.edu/Reports/2015/Sharing-Clinical-Trial-Data.aspx>

22. Starr, J. *et al.* Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Comput. Sci.* **1,** e1 (2015).

23. Altman, D. G. & Moher, D. Declaration of transparency for each research article. *BMJ* **347,** f4796 (2013).

24. Laine, C., Goodman, S. N., Griswold, M. E. & Sox, H. C. Reproducible Research: Moving toward Research the Public Can Really Trust. *Ann. Intern. Med.* **146,** 450–453 (2007).

25. The YODA Project | Policies & Procedures to Guide External Investigator Access to Clinical Trial Data. at <http://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data>

26. Krumholz, H. M. Why data sharing should be the expected norm. *BMJ* **350,** h599 (2015).

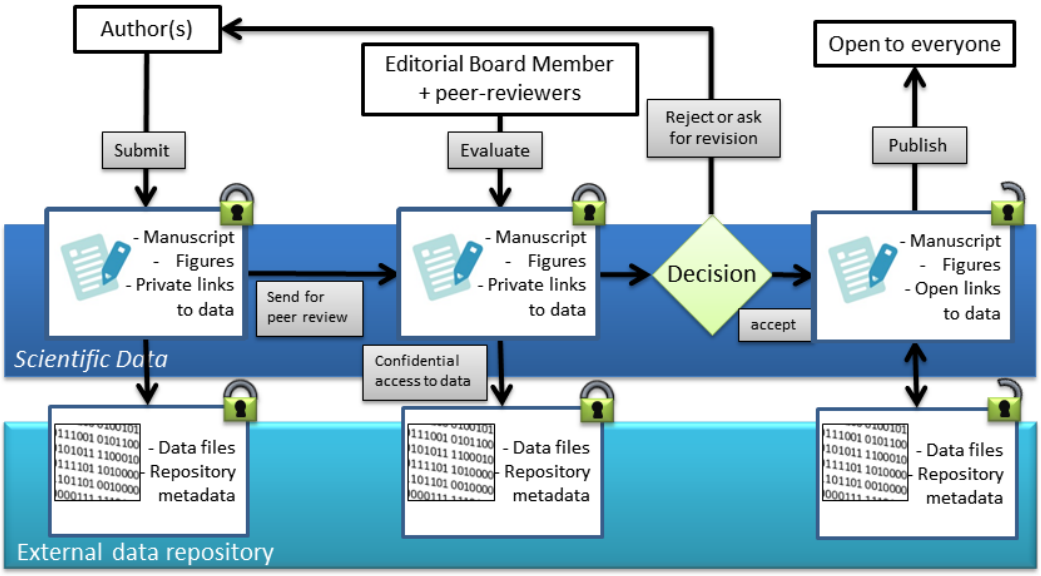
Author(s)
Editorial Board Member + peer-reviewers
Open to everyone
Submit
Evaluate
Reject or ask for revision
Publish
- Manuscript - Figures - Private links to data
Send for peer review
- Manuscript - Figures - Private links to data
Decision
accept
- Manuscript - Figures - Open links to data
Scientific Data
Confidential access to data
- Data files - Repository metadata
- Data files - Repository metadata
- Data files - Repository metadata
External data repository

Figure 1: Overview of standard *Scientific Data* editorial workflow for non-confidential datasets

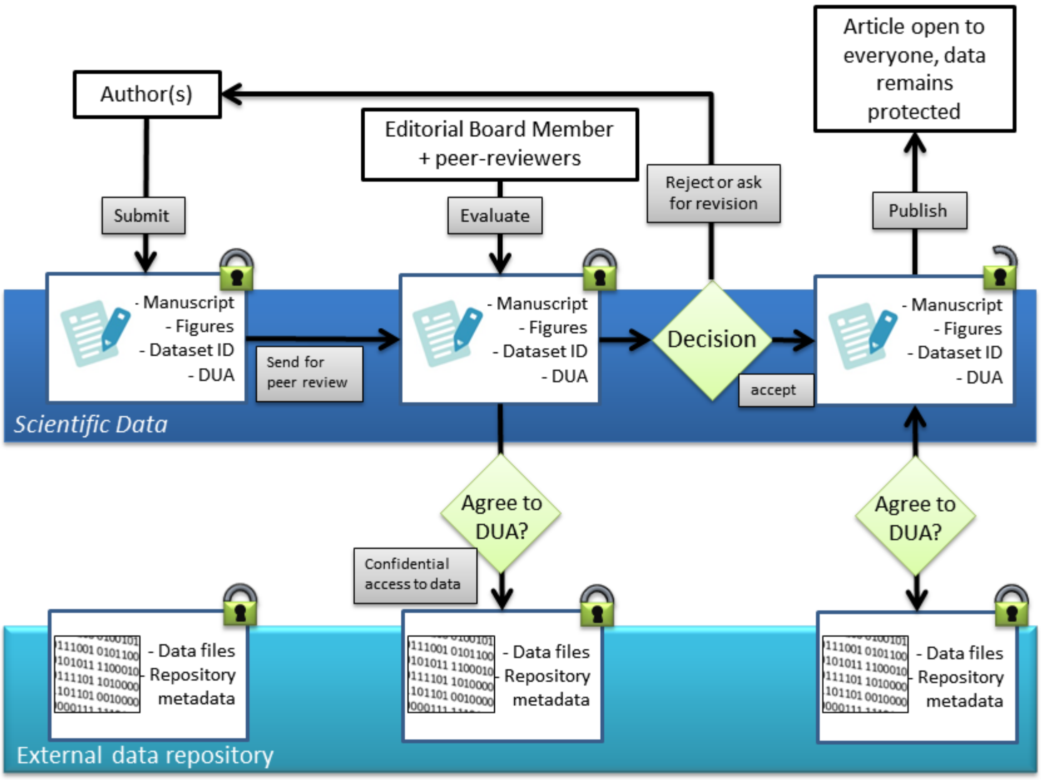

Figure 2: Overview of likely editorial workflow to accommodate peer review and publication of clinical Data Descriptors