



# A beginner's guide to data stewardship and data sharing

Marcel P. Dijkers<sup>1</sup>

Received: 23 September 2018 / Revised: 28 November 2018 / Accepted: 3 December 2018 / Published online: 5 February 2019  
© International Spinal Cord Society 2019

## Abstract

**Study design** A narrative review of principles, benefits and disadvantages, as well as methods of research data sharing.

**Objectives** To assist prospective Spinal Cord authors and others with understanding and implementing data sharing, so that various benefits of such sharing can accrue to all spinal cord injury research stakeholders.

**Setting** International.

**Methods** The medical research and health care services literature was reviewed nonsystematically for relevant articles, and web sites were explored for information and services offered by various pertinent organizations.

**Results** Grant makers, professional organizations, research journals, publishers, and other entities in the research field increasingly stress the ethics as well as societal and practical benefits of data sharing, and require researchers to do so within a reasonable time after data collection ends. Sharing data, retrospectively, generally requires much time and resources, but when a data management plan is part of a research proposal from the start, costs are limited, and grant makers allow these costs to be part of a budget. There are many organizations that offer information on or even assist with preparing data for sharing and actual deposit in a data repository.

**Conclusions** The requirement of data sharing is not likely to go away, and researchers interested in submitting their reports to Spinal Cord would do well to familiarize themselves with the myriad practical issues involved in preparing data for sharing.

## Introduction

Similar to many other journals in biomedicine and health care, *Spinal Cord* strongly encourages all authors submitting a manuscript to the journal to make their data available upon acceptance of the paper. (See the editorial in this issue[1]). This paper offers an overview of the main issues involved with such open data, including the why, when and how of sharing of data. The focus of this article is helping researchers who want to share their data. Some of this information (e.g., on the re3data website, which is a resource for finding data repositories [2]) may also be helpful to users of such data, as are many papers in the literature, e.g., [3–7].

The term “open data” is used to refer to the “raw” or minimally processed data created by or used in the research underlying a report in the scientific literature, and made

available, with or without conditions for commercial and noncommercial use, to other investigators and even to nonscholars, to use, reuse and redistribute, free or at nominal cost. Often, the data underlying the findings reported in a paper are just a selection of all data collected in a particular research project [8], and the idea underlying “open data” is to publish all data amassed, not just the subset used in a particular report, so that fellow scientists who want to validate the analysis can convince themselves that no cherry picking (data mining, p-hacking) took place. Also, data per se are useless without their metadata: information on how the data were collected, transformed, and organized in the data file shared with others. (Metadata means data about data). When the term “open data” is used here, it refers to all data generated by a research project, together with all their metadata, unless otherwise specified.

The concept of “open data” is relatively new; the first publications in PubMed using this term date from the early 2000s. It is part of a push (culture shift) toward “open science”—the idea that all science should be conducted transparently [9, 10], with all tools (concepts, protocols, code, data, etc.) readily accessible, and all benefits resulting from its findings potentially available to

✉ Marcel P. Dijkers  
marcellinus.p.dijkers@wayne.edu

<sup>1</sup> Department of Physical Medicine and Rehabilitation, Wayne State University, Detroit, MI, USA

every human [11, 12]. “Open access”, the notion that scientific publications should be available free of charge to any reader (with the cost of publishing borne by the investigator or the sponsor of the research), is a connected idea. The idea of open data has been worked out most consistently in the FAIR principles, a set of research community-generated guiding standards that are concise and independent from the area of science one is active in, and which apply equally to metadata and to data. Together, the FAIR principles “facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows” [13]. The four facets of FAIR are that data should be, for anyone:

- **Findable**—through rich metadata which include a unique and permanent identifier for the dataset
- **Accessible**—retrievable using an open, free communications protocol, at a price that is at most nominal
- **Interoperable**—can be linked through references to other (meta)data and use of formal, broadly applicable language for knowledge representation
- **Reusable**—released with provenance and a clear and accessible data usage language [13].

When one considers that medical and health research data consist not just of digital questionnaire responses and clinical trial data, but also include images, auditory tracings, genetic sequences, and many other forms of information, it becomes

clear that implementing the FAIR principles across all areas of research is not simple. Box 1 offers a sampling of organizations that are involved in developing standards and systems that make data exchange and reuse simple.

The claim can be made that with open data, science is returning to its roots: in the 17th century individual scientists exchanged their observations as raw data with others, initially in letters and then in the first journals [14]. The small circle of gentleman-scholars back then has expanded into a world-wide cadre of millions of professional scientists and clinician–scientists, who collect information that is quantitative rather than qualitative, and produced with often sophisticated instruments rather than the naked eye. Data have moved from laboratory notebooks to punch cards to electronic files, which are easily shared with others, on magnetic data storage tapes and disks, and now over internet connections. But the most important driver of open data may be that most branches of science have become or are becoming more data-intensive, producing “Big data” that cannot be shared in a simple table in a paper that gives all information by case number [15]. Correctly analyzing these mountains of data requires sophisticated software to find patterns [16]; the papers that may result cannot but offer a very small quantitative summary derived over many steps of analysis of thousands or millions of data points. The quantity of data (bytes) that Darwin collected over a lifetime now is produced by a magnetic resonance imaging machine or a high-energy particle telescope in seconds. Would Darwin have endorsed open data?

#### Box 1 Helpful data stewardship and data sharing resources for beginners

Resources marked with \* may charge fees for some or all their services or products.

##### (1) Australian National Data Service (ANDS)

A program supported by the Australian government, to deliver national services to support data discovery, connection, publishing, sharing, use and reuse. Available for free on its website is a wealth of materials for building the data skills and capacity of researchers and research organizations, Australian or not. A sampling of their guides:

- Checklist for assessing IT Infrastructure capability for data management
- Copyright, data and licensing
- Curation continuum
- Data management plans
- Data sharing considerations for Human Research Ethics Committees (HRECs)
- Data storage
- Deidentification
- Digital Object Identifier (DOI) system for research data
- Metadata
- Publishing and sharing sensitive data

##### (2) Clinical Data Interchange Standards Consortium (CDISC)

“The CDISC mission is to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.” CDISC develops standards supporting clinical and nonclinical research processes, including models, domains and specifications for data representation. Training around the implementation of the standards is available in the form of courses and webinars.

**(3) Consortium of European Social Science Data Archives (CESSDA)**

The Consortium offers in-person training courses, and various materials on its website, including:

- Tutorial: Elements of the ingest and access/dissemination workflow
- Research data management community training
- Webinar access policies and usage regulations: licenses

**(4) Creative Commons (CC)**

CC is a nonprofit aiming to expand the type and number of creative works available for others to build upon (including research data and their metadata) legally and to share. It publishes a number of copyright-licenses known as Creative Commons licenses, available free of charge, which allow data creators to communicate which rights they reserve, and which rights they waive for the benefit of others.

**(5) Dataverse**

Dataverse is an “open source web application to share, preserve, cite, explore, and analyze research data.” It facilitates setting up web sites for sharing research data, and offers source codes and accompanying guides for doing so.

**(6) Digital Curation Centre (DDC)\***

DDC is “an internationally-recognized centre of expertise in digital curation with a focus on building capability and skills for research data management.” It offers an extensive list of resources, training, and technical assistance.

**(7) Dryad\***

Dryad is a curated repository that accepts datasets and makes the data (of a wide diversity of formats) underlying scientific publications discoverable, freely reusable, and citable. Dryad is searchable.

**(8) eResearch Persistent Identifiers Consortium (ePIC)**

ePIC offers software and information to create persistent identifiers for research products, including datasets, that are alternatives to DOIs.

**(9) figshare\***

figshare is a repository where researchers can share their research outputs, including figures, datasets, images, and videos, and about any other format, after uploading the data or information in a few simple steps. The site is searchable so that potential users can identify useful datasets among the millions now contributed.

**(10) Go FAIR**

The organization developed to support implementation of the FAIR principles. Some of the resources available are:

- Go FAIR metadata
- Go FAIR tools
- FAIRification process
- Research data management starter kit
- Webinars
- Metrics
- Videos
- Glossary

**(11) International DOI Foundation (IDF)\***

IDF is a not-for-profit membership organization that governs the federation of Registration Agencies providing Digital Object Identifier (DOI) services and registration. The DOI system provides a technical and social infrastructure for the registration and use of persistent interoperable identifiers, called DOIs, for use on digital networks. The website provides factsheets, links to registration agencies, and a wealth of other information about obtaining a DOI.

**(12) International Spinal Cord Society (ISCoS) International SCI Data Sets**

The ISCoS data sets offer a Core Data Set, as well as system/ problem/ outcome Basic Data sets and Expanded Data Sets to describe patients/ subjects with SCI and their functioning in a standardized way that optimizes sharing of data.

**(13) Inter-University Consortium for Political and Social Science Research (ICPSR)\***

A data repository specializing in social science data. Offers a wealth of information on its website, as well as summer training programs. Information on the website includes:

- Preparing data for archiving
- Guide to archiving social science data for institutional repositories
- Curating data at ICPSR
- Trusted digital repositories
- Digital preservation policies and planning at ICPSR
- Data management plans
- Guidelines for US Office of Science and Technology Policy data access plan

**(14) MANTRA**

MANTRA is a free, online course on management of digital data, directed at postgraduate students, early career researchers, and information professionals, developed by staff of **Edina (15)**, a program of the University of Edinburgh focused on development of tools to work with data.

**(16) National Institute for Neurological Disorders and Stroke Common Data Elements for SCI (NINDS CDEs)**

The SCI CDEs are an expansion of the ISCoS data sets, providing new data sets and adding metadata elements to the existing ISCoS descriptions so as to enhance data sharing. The website offers:

- CDEs—classified as Core, Supplemental-Highly Recommended, Supplemental, and Exploratory
- CRF (Case Report Forms) Modules—logically organized CDEs for data collection
- Guidelines—to provide further information about the CDEs
- A study start-up tutorial

**(17) National Institutes of Health Data Sharing (NIH)**

This page offers the NIH policy on data sharing, but also some resources such as:

- NIH data sharing policy and implementation guidance
- Frequently asked questions on data sharing
- Data sharing workbook
- Other data sharing documents and resources

**(18) Open Data Commons (ODC)**

ODC is a project of **Open Knowledge International (19)**, and specializes in making available legal tools for open data. It does that by publishing licenses, and its website offers a variety of related information, including:

- Introduction to open data
- Open definition for data
- Quick guide to making data open
- Open data handbook

**(20) Registry of Research Data Repositories (re3data)**

re3data is nothing more (or less) than a registry of data registries—over 2000 as of this writing. As such, it enables researchers to quickly find the registry where their data might be best stored. For each registry, re3data offers subject(s) covered, content type(s) (e.g., standard office documents, images, scientific and statistical data formats, plain text), country and Uniform Resource Locator (URL), and a synopsis of contributors to and purpose of the registry.

**(21) Research Data Alliance (RDA)**

RDA has a number of Working and Interest Groups (in the areas of Agriculture, Biomedical Sciences, Chemistry, Digital Humanities, Librarianship/ Archival Science and Information Science, and Linguistics) to develop and adopt infrastructure that promotes data-sharing and data-driven research. Its major outputs are Standards and Recommendations.

**(22) Springer Nature Research Data Support\***

Assistance for researchers publishing in Springer Nature journals, about the journals' policies on open data, selecting repositories, preparing data, etc. Information on the website includes:

- Research data helpdesk
- Recommended repositories
- Journal policies and services

**(23) UK Data Service**

The UK Data Service offers various sets of data from UK government data collection efforts, and other social and economic data, but also:

- Supports for policy-relevant research
- Guidance and training for the development of skills in data use
- Supports for the development of best practice data preservation and sharing standards
- Shared expertise with international data providers to remove barriers to accessing data

**(24) Vivli**

Vivli is a Harvard University-affiliated nonprofit organization offering a global clinical research data sharing platform, intermediary between data producers and data users as well as the larger research community.

**(25) Wellcome Trust Data Sharing**

The Trust offers a number of resources on data sharing (e.g., the role of the European General Data Protection Regulation [GDPR] in research), including the following reports

- Building and sustaining data infrastructures
  - Interoperability standards
  - Developing skills for managing research data and software in open research
1. ANDS. Australian National Data Service. <https://www.ands.org.au/>.
  2. CDISC. CDISC Standards in the Clinical Research Process. <https://www.cdisc.org/standards>.
  3. Consortium of European Social Science Data Archives—European Research Infrastructure Consortium. <https://www.rd-alliance.org/consortium-european-social-science-data-archives>.
  4. Commons C. When we share, everyone wins: @creativecommons.s <https://creativecommons.org/>.
  5. The Dataverse Project. <https://dataverse.org/>.
  6. DCC. Digital Curation Centre | because good research needs good data. <http://www.dcc.ac.uk/>.
  7. Dryad. Dryad Digital Repository. <https://www.datadryad.org/>.

8. ePIC. Persistent Identifiers for eResearch. <https://www.pidconsortium.eu/>.
9. figshare. figshare - credit for all your research. <https://figshare.com/>.
10. FAIR Principles - GO FAIR. <https://www.go-fair.org/fair-principles/>.
11. doi. Digital Object Identifier System. <http://www.doi.org/>.
12. ISCOS. International SCI Data Sets. <https://www.iscos.org.uk/international-sci-data-sets>.
13. ICPSR. Available from: <https://www.icpsr.umich.edu/icpsrweb/>.
14. Research Data MANTRA. <https://mantra.edina.ac.uk/>.
15. EDINA. <https://edina.ac.uk/>.
16. NINDS. NINDS Common Data Elements. <https://www.commondataelements.ninds.nih.gov/#page=Default>.
17. NIH Data Sharing Policy. Available from: [https://grants.nih.gov/grants/policy/data\\_sharing/](https://grants.nih.gov/grants/policy/data_sharing/).
18. Open Data Commons <https://opendatacommons.org/>.
19. Open Knowledge International. <https://okfn.org>.
20. re3data. [www.re3data.org/](http://www.re3data.org/).
21. Research Data Research Data Sharing without barriers. 2018.
22. Research Data Support | Authors | Springer Nature. [https://www.springernature.com/gp/authors/research-data-policy/?utm\\_source=Other\\_digital&utm\\_medium=Other\\_digital&utm\\_content=MatAst-Nature-SNRDS\\_home-Multidisciplinary-Global&utm\\_campaign=SCIDATA\\_SUB\\_SNRDS\\_home](https://www.springernature.com/gp/authors/research-data-policy/?utm_source=Other_digital&utm_medium=Other_digital&utm_content=MatAst-Nature-SNRDS_home-Multidisciplinary-Global&utm_campaign=SCIDATA_SUB_SNRDS_home).
23. UK Data Service. <https://www.ukdataservice.ac.uk/>.
24. Vivli - Center for Global Clinical Research Data. <https://vivli.org/>.
25. Wellcome Trust. Our new policy on sharing research data: what it means for you. <https://wellcome.ac.uk/news/our-new-policy-sharing-research-data-what-it-means-you>.

## The why of open data

Proponents of open data have pointed out benefits to society as a whole, and for scientists more specifically; these of course are related issues. From a societal point of view, the major claim is that data created with public (government) or semipublic (foundations large and small) funding should be easily and openly publicly available. There is no need to spend taxpayer money on collecting new data when an existing dataset can be used to answer (most of) a research question. One might also argue that in medical and behavioral research, study volunteers have given time and effort, sometimes even risked their health and well-being, to help scientists obtain their data; their sacrifices are properly compensated by participant fees, but also by reuse of the information they “produced”. Lastly, the claim is made that the free availability of data speeds up discovery and creation of products that have the potential to benefit all.

### Benefits to science

The benefits to other scientists and science as a whole of sharing one's data are stated to include the following [14, 17]:

- Others can access data to verify and/or expand the published analysis performed by the researcher producing the data. This verification goes beyond what can be done by peer reviewers, who typically only appraise the summarized data and the author's interpretations of these data. (Note that some journals now require making the dataset underlying a paper available to the editor/peer reviewer, for them to be better able to evaluate a manuscript and the claims made in it. A few journals have gone so far as to made acceptance of a paper

conditional on the data being openly available to everyone).

- Similarly, access to the entire dataset makes it possible for editors, peer reviewers and others to determine whether a (to be) published report is based on selective outcome or subgroup reporting, or even worse, selective publishing, which big pharma has been accused of doing.
- In these years of a reproducibility crisis, where results of many scientific studies are difficult or impossible to replicate by others [10], access to the raw data is, of course, also a check on data fabrication and falsification, although it must be admitted that a skillful forger can create data that can be hard to distinguish from the real thing.
- Scientists who have access to the data can compare them with their own or yet others' data, and start a debate on issues of data collection, analysis and interpretation based on information that is wider than the (always subjective) selection and interpretation offered in the article(s) the data creators wrote. Especially in the case of megatrials, access to the unique data is a sine qua non, as it is not likely that funding for a parallel multisite, multiyear trial will ever be made available [18, 19].
- Others can add their own data to create a richer dataset (more cases, or more variables on the same cases; see FITBIR below) or combine published datasets from multiple studies to answer their questions using many more cases—as in, for instance, individual patient data meta-analysis (IPDMA).
- Other scientists can use the data to answer completely new questions, sometimes based on secondary analysis for a (preliminary) hypothesis, sometimes based on data mining. Even if an available dataset might not answer their questions, it may at least enable them to assess

whether new data collection of their own is likely to be able to answer their questions. Similarly, access to data can assist in designing new studies—e.g., in performing a power analysis. Thus, open data can help to reduce the waste said to characterize preclinical and clinical research [20, 21].

- And not least, sharing data through a public repository keeps them from being lost even as the scientist who produced them changes jobs, computers, and storage media [22].

### Benefits to individual scientists

If you share the values and principles that underlie the open data movement, the satisfaction of acting in accordance with your principles is a good enough reason to make your research data available. However, sharing takes time and money, and other incentives may be needed. Among those can be named:

- Satisfying obligations imposed by a grant maker or other bodies involved in administering and regulating your research. Many grant makers nowadays are requiring that scientists submit, as a part of their research grant applications, a data management plan that includes statements on how data will be shared, when and with whom. Some funders check on the implementation of the promises made, and some now, in making decisions on a new grant, take into account actual follow-through on sharing of data produced in a previous grant-supported project. And when the stream of grant funding is endangered by such stipulations, universities and research institutes step in, forcing the Principal Investigator (PI) into data sharing, often offering mechanisms for doing so.
- A dataset made publicly available may count as an additional research product, beyond the traditional research paper, and other, more modern, derivatives of data. In some areas of science, a paper that essentially provides the metadata for an available database (a “data paper” or “data descriptor paper”) counts as a legitimate publication [23]. Similarly, the creation of a dataset for the benefit of the discipline as a whole, without a paper describing analyses and findings, may count as a valid research product. This all represents an expansion of scientific productivity metrics beyond authorship counts [24]. It must be admitted that, as of now, there is no clear consensus how the publishing of datasets and other nontraditional products is to be considered in the evaluation and promotion of scientists [12], and it presumably will be some time before individual disciplines work out the rules.

As indicated below, open data that are deposited in the better repositories receive a digital object identifier (DOI), which DOI can be provided in a traditional research paper and otherwise disseminated. The use of the dataset can, in turn, by way of the DOI lead other researchers to the papers of the database creator. This could lead to more citations of her papers, which may enhance current indicators of scientific productivity, the h-index and similar derivatives from impact factors [11]. Piwowar and Vision reported that “studies have found that papers with publicly available datasets receive a higher number of citations than similar studies without available data” [25, p. 1] and confirmed this in their own study, which controlled for multiple factors affecting citation rates.

- In some scientific collaboratives, use by outsiders of the dataset produced by the collaborating scientists requires these outsiders to make the data creators a co-author or contributor on whatever papers result from the secondary data analysis, adding directly to the resume of the data producers [26]. The rules of authorship promulgated by such organizations as the International Committee of Medical Journal Editors (ICMJE) [27] may not necessarily condone such “passive authorships”, but there may be opportunities (as discussed below) to make these partnerships more active.
- Unless that what the fellow scientists, attracted by a published dataset, propose is completely different from the additional analyses the data creator wants to do, publication of data may result in collaborators who are willing to do the heavy lifting. The scientists collaborating in the US Spinal Cord Injury National Data Base (SCI NDB), which contains “open data” (see below), offer outsiders the opportunity to be given a mentor who is familiar with the database. If a match of interests between outsider and mentor can be found, a very productive partnership may ensue. The concern is sometimes expressed that even with rich metadata, a secondary analyst lacks the detailed knowledge that the original investigators developed [28, 29]; mentorship may be the answer.

There is some research in this area. In a 2011 study by Tenopir et al., 85% of investigators were interested in using other researchers’ data—if easily accessible; 74% thought that their own data could be used for new purposes; but only 36% considered those data easily accessible [30].

### Costs to society, science, and scientists

Often issues are brought up that would seem to argue against public release of all, or at least most, research data.

The first among these is that the data creator “owns” the data, and as such can treat them any way he wants, including destroying them. That may be true for research done by private companies, but (semi)public funding changes the arithmetic. And: generally, a grant is made to a university or other research institution on behalf of a PI—which likely means that the university “owns” the data, not the PI.

A more serious issue, especially with respect to human subjects research, is the need to protect the private and sometimes sensitive information that subjects have allowed the research staff to collect. Public knowledge might result in stigma, exposure to criminal prosecution, or economic disadvantages for the subjects. Similar protection of data sometimes is warranted for social groups, businesses, and at-risk cultural remains and ecosystems. The counter-argument here is that in most instances the data (the cases) can be de-identified, so as to reduce or even completely eliminate the risk. In situations where that is not possible, data can be made available to other researchers under strict conditions that protect the research participants (see below). There has been quite some debate whether re-identification is possible to circumvent the safety shields put in place; the consensus seems to be that with use of up-to-date “strong” deidentification techniques, data of high quality (i.e., informativeness) can be released that provide adequate safety for individual privacy [31].

Without any doubt, preparing the data and especially the metadata that make them useful to others takes time. The time required may be limited if a data management plan is created at project start and followed consistently (see below), but there is no doubt that there is a scientist and support staff time involvement, as well as sometimes other costs. The good thing is that the funding agencies that require a data management plan and the publication of open data are willing to put their money where their mouth is: the costs involved can be put into the grant budget.

Another cost often quoted by those opposed to open data is the loss of opportunities to derive additional papers from the data—the thinking being that the moment data are out there in public storage, the creator of the dataset will be scooped. However, the fact that others **can** analyze one's data is no guarantee that they **will**—the current climate of science emphasizes originality over confirmation and replication, and most scientists will be more interested in creating and analyzing their own data. The fear of “research parasites” [32] likely is pathologic [22, 33]. For instance, the SCI NDB, a longitudinal study of the circumstances, treatment and lifetime consequences of SCI that started in 1973 and has information on over 30,000 cases [34] has received less than 10 requests for access to the data since that opportunity was first opened up a decade ago (personal

communication, Yuying Chen, 2018). An additional “protection” of the rights of the data creator is the fact that an embargo period is acceptable—most grant makers allow the researcher about a year after the funding period ends to wrap up publishing—which certainly is sufficient for RCTs and similar short-term studies. For truly longitudinal studies (e.g., the SCI NDB) where there is no “end” to data collection, another solution needs to be found. The SCI NDB has a 5-year embargo (outside researchers cannot obtain the most recent 5 years' worth of data collected), which can be waived, however, and routinely is.

Data dredging by data borrowers is another reason that has been put forward for not sharing data. (As if data creators are any less skillful at p-hacking than secondary analysts!) Hlatky and Januzzi offer a set of “Proposed additions to journal instructions to authors for manuscripts derived from shared datasets” that addresses, among other issues, being open about the degree of dredging [35, p. 1984]. Some have argued that secondary analyses published in the literature may, by errors of omission or commission, or because of equally valid alternative analysis methods, come to conclusions that are contrary to the ones reached by the data creators in their own report [17]. That is certainly possible, but is no reason to not make data available. It is not a new danger in research—incompatible findings and disagreements on the proper analysis and interpretation of data are common. Science is one of the few human enterprises where peer review is a sine qua non of acceptance of one's products, and open data just extend the risk that traditionally was offered by peer reviewers and letters to the editor, because more information to find fault with is now available. Even if errors are detected, the scientist who made those mistakes experiences no negative consequences, if it is clear that she acted in good faith. Different approaches and interpretations contribute to scientific progress, and trying one's best to avoid errors does so even more [22].

The same might be said with respect to the ultimate humiliation: a secondary analyst comes up with a spectacular finding that opens up grand new avenues of research leading to a Nobel prize—which avenues the data creator missed. Well, Newton may have missed things that Einstein picked up on. That's the nature of our craft.

## Endorsements

Because of the multiple benefits, to science and scientists themselves and to society at large, and the limited or surmountable counterarguments, a number of grant funding agencies, publishers, journals, and organizations of scientists are calling for implementation of open data in some form or another. The following is a sampling:

- Public and private funding agencies. For instance, in the United States the National Institutes of Health, the world largest sponsor of biomedical research, has put in place policies requiring all (large) research projects to make their data publicly available by 1 year after grant funding end [36]. Similar policies are now in place at the US National Science Foundation [37] and, in the private sector, at the Gates Foundation [38]. In Great Britain, the Wellcome Trust, a major private sector funder, has worked with other funders of medical research to develop standards and policies requiring open access [39]. The Canadian Institutes of Health Research (CIHR) also strongly support open data [40], while the Organization for Economic Co-operation and Development (OECD) is on record with “Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based” [41, p. 15].
- Journal editors and publishers. ICMJE, an influential group that often is in the vanguard of new research policy trends, requires (starting mid 2018) that the data of all clinical trials be made publicly available, within 1 year of publication [42]. In this instance, ICMJE is lagging some journals. For instance, the large and well-respected open-access publisher Public Library of Science (PLOS) already requires that for every paper published in a PLOS journal, all data are made available, “with rare exceptions” [43]. The publisher of Spinal Cord, Springer Nature, also has been a forerunner, and offers a number of services to scientists who need assistance in publishing their data [44] (see Box 1). Spinal Cord itself has the following on its website in the “For authors” section: “Research Data Policy. We strongly encourage that all datasets on which the conclusions of the paper rely should be available to readers. We encourage authors to ensure that their datasets are either deposited in publicly available repositories (where available and appropriate) or presented in the main manuscript or additional supporting files whenever possible. If a public repository does not exist, the information must be made available to editors and referees at submission and to readers promptly upon request. Any restrictions on material availability or other relevant information must be disclosed in the manuscript’s Methods section and should include details of how materials and information may be obtained” [45].
- Organizations of scientists. The Institute of Medicine (IOM), a think tank with much influence on the policies of various elements of the US government, in 2015 published a report “Sharing Clinical Trial Data. Maximizing Benefits, Minimizing Risk”, in which it

endorsed open data for clinical trials [46]. The American Psychological Association’s (APA) Ethical Guidelines for Research state (8.14) that data sharing for verification is required, but (as of now) the APA does not require data sharing for secondary analyses [47]. Not all psychologists concur with that timid stance [48].

## How to share one’s data

In a 2017 survey of over 7700 scientists conducted by Springer Nature, 76% gave high importance to making their data discoverable [49]. However, they also identified the following challenges to doing so:

- Uncertainty about copyright and licensing of data (37%)
- Lack of knowledge as to which data repository to use (33%)
- Lack of time to deposit data in a useful way (26%)
- Costs involved in sharing data (19%)

This section addresses a number of these issues, mostly by pointing out resources (Box 1) available to busy (clinician-)researchers with always limited funding to make “open data” doable.

The easiest road to “publishing” data, consisting of the data themselves and the metadata, is to plan for it from the beginning of a project, which generally is the submission of a grant proposal, including development of a data management plan that lays out at which point or points of time what specific data will be released to whom, in what format and under what conditions, if any. Whether one considers statistical analysis code and other scientist-created software part of the metadata or not, making them available will make the data per se more useful for others. Staff of the Inter-University Consortium for Political and Social Science Research, a data repository with over 50 years of experience, plausibly claim that if data are managed properly from the start of a project, with active data curation and documentation using software specialized for that job, the need for extensive efforts just before sharing is greatly reduced [50].

Grant proposal preparation is also the time to contact one’s Institutional Review Board (IRB), Human Research Ethics Committee (HREC) or equivalent, if applicable, with a request for permission to release a dataset with de-identified data. Many issues regarding the release of confidential private data of individuals can be avoided by proposing an informed consent document in which subjects approve release of their de-identified data. Requesting such permission generally has not been found to reduce the percentage of participants who enroll in the research [51, 52], but if such seems likely, an opt-out for the data

release can be offered, which will affect only the size of the open data set distributed.

Prospective development of an open data release also makes it possible to incorporate clinical data interchange standards [53] and even common data elements (CDEs) into one's data [54]. Using these selected information elements has several advantages: the developers of the CDEs likely have created a full description of each data element, complete with formats to be used in coding and storage—potentially tremendously reducing the amount of such work the investigator and her staff need to do (see below). Secondly, scientists using the open data are likely to have familiarity with the CDEs, reducing their learning curve. Maybe most importantly, multiple datasets all using the same CDEs can be merged [34] without weeks of recoding variables to make them combinable, as is not unusual for IPDMAs [55]. The specific electronic format (s) in which data are prepared and distributed depend on a number of factors, among which discipline traditions may be foremost: many scientific fields have developed standards that make the actual distribution and use of open data almost routine.

All the issues mentioned thus far (especially the obligation to share data) and listed below (deidentification, data licensing, and platforms for distribution) become even more complex when investigators at multiple institutions, possibly in different jurisdictions, contribute to a large dataset. An Open Data subcommittee of the steering committee is needed to work its way through the alternatives available and propose solutions, and every clinician thinking of contributing data must consider her agreement with those solutions before providing data on her first case.

Turning data collected in the past, without the benefit of a data management plan, into open data is much more difficult. One estimate is that it takes a week to prepare a trial dataset and its dictionary metadata for sharing [56]. Even scientists who pride themselves on their meticulous documentation of all data collection, processing and analysis steps may be disappointed to find out how much they rely on memory (or casual notes on slips of paper in a folder) to guide them in taking next steps and writing a formal report. For metadata, memory is no good, and complete descriptions of every data element, every step in processing, and each file storage element need to be produced. The “Data dictionary” for the SCI NDB runs 687 pages, and even at that heft is incomplete, in that it refers to other documents. Most projects will be simpler, in covering fewer centers, variables or time points at which different types of information are collected, but still will need many pages of detail to provide a secondary analyst with all the information she needs to truly understand the nature, value, completeness, and precision of the open data she wants to use.

## Deidentification

Both prospectively and retrospectively collected data will need to be de-identified—that is, stripped of information that can give away the private information of individuals, social groups and other entities that need protection. Under more stringent standards, information that allows linking to other datasets from which identification is possible also needs to be scrubbed [57]. Sometimes deidentification can be achieved by recoding information—e.g., when census block codes are combined into a county code. The nature of what has to be eliminated or recoded depends on local laws and customs [58, 59], and limited generic information can be provided here; investigators should discuss these issues with the local IRB/HREC staff, grant funders, their institution's legal and other administrative staff, and government offices charged with the protection of patients and other vulnerable groups. Most have extensive guidance, training, and application materials on their website. Hrynaszkiewicz et al. [60] provide useful information on anonymization, and provide a list of 28 potential patient identifiers that need to be removed, which list includes the 18 items covered by the Health Insurance Portability and Accountability Act (HIPAA) in the US.

## Options for sharing one's data

The conditions under which the data are made “open” to a degree determine the extent of deidentification needed. There likely are dozens of options, but the following points along a continuum may be given:

1. The data are published in the paper reporting on the research (e.g., a small table listing all variables by case number for a case series), or as supplemental digital content (SDC) for the paper, on the website of the journal. The data format in SDC should be a common file format (e.g., Excel), and SDC is limited to fairly small datasets. The disadvantage of this method is that no DOI with all its benefits is created for the dataset per se; the advantage is that the data creator never again has to attend to any issues around data release. Spinal Cord accepts SDC, including these small datasets.
2. The data are submitted to a data storage facility such as Figshare [61] or Dryad [62] from which everyone can download them, together with the metadata, without having to ask anyone for permission. (This approach is also known as an unsupervised data archive). Many universities and other scientific facilities offer the same opportunities [63–66]. This option allows for submission of datasets of almost unlimited size, and most repositories issue DOIs; there

is no difference with the previous option in terms of the need for deidentification.

3. An entity such as the National Spinal Cord Injury Statistical Center (which manages the SCI NDB) serves as the curator of the data. A scientist who desires to use them must submit a proposal stating what specific data she wants to use, and for what purposes. The proposal and the scientist are vetted for their quality, and if approved, a dataset limited to what was requested or approved is released. (This approach is also known as a supervised data archive). The advantage of this model is that data release can be made conditional on a binding contract (a data use agreement) specifying what the scientist is and is not allowed to do with the data—for instance, reidentifying the cases, redistributing the data, and any analysis not contained in the proposal. In addition, some curators reserve the right to review any manuscript written, prior to its submission to a journal, to make sure there is no misunderstanding of the data or misrepresentation of those who collected them. Data use agreements also may allow for release of nondeidentified data, which may make data linking to other relevant datasets possible.
4. An interesting option is offered by the U.S. Federal Interagency Traumatic Brain Injury Research Informatics system (FITBIR) [67]. FITBIR functions as a supervised data archive, accepting datasets from multiple investigators, and making them available to others who have been approved and have signed a data use agreement. However, based on a Global Unique Identifier (GUID) (a unique permanent identification number that is derived from personal information such as birthdate, but cannot be reverse engineered to produce that birthdate), researchers can contribute data that can be linked to data on the same person but created as part of a different research project, possibly across the country.
5. The most stringent conditions and best protection of individual privacy are offered by curators who do not release a copy of the database. Those scientists interested in “using” the data can submit a request for specific tabulations to be prepared by the curator’s statistical staff, which summary data are then supplied upon approval of the request. (This is known as a supervised data enclave). Sometimes the entire process is automated—the user specifies variable names, and the computer produces the tabulations. (Also termed an unsupervised public data enclave, if anyone can submit requests, or an unsupervised private data enclave, if that right is limited to prequalified and registered users.)
6. A last option is “data available from the author upon reasonable request”, which offers the advantage that the

data creator stays absolutely in charge of his data. Otherwise, there seem to be only disadvantages all around: he has to personally review and fulfill each request that comes in, can apply very idiosyncratic rules as to who or what is approved, and there is no certainty that the dataset will survive, in the short term let alone the long term. Even before the demise of the data (or of the researcher), there may be problems with this method of open data distribution, for other scientists: Wicherts et al. [68] report that 73% of 149 author teams never complied with a request for data stated to be “available on request”.

Availability of data repositories for SCI data figured prominently among the initiatives to facilitate SCI research collaboration listed by 213 SCI clinicians and researchers in a 2018 study [69], but among the barriers, intellectual property and privacy issues were named. As of this writing, there are no repositories specialized in SCI data, whether preclinical, clinical or community functioning-focused [70]. Some preliminary steps that will make exchange of truly useful data sets feasible have been completed, though [71]. However, the current unavailability of a SCI-focused repository should not stop SCI researchers from sharing their datasets as open data; options 1, 2, and 6 above are open to them.

The Digital Curation Centre [72] has an extensive checklist “Where to keep research data” that can help a researcher determine the best repository for her data. Burton et al. provide a technical discussion of criteria for “data safe havens”, a term often used for repositories of sensitive data [73].

### Copyright issues

In addition to privacy law, the rights and obligations that apply to research data may include intellectual property rights (copyrights, database rights, and patent rights), national security law (for secret research conducted within and outside of government), trade secrets, and contractual agreements, both as incorporated in research grants made by governmental or private entities, and research contracts with industry [6]. There is too much variety in legal standards and specifics of signed agreements that it makes sense to address the issues here—with one exception: copyright.

Except maybe for a few jurisdictions (Australia keeps coming up), the *facts* of nature cannot be copyrighted, but the *format* in which these facts are arranged by scientists can. “... numeric values at the item level are ‘uncopyrightable’ data elements in a data set in most of the world” [6] p<sup>423</sup>. In other words, if Mendeleev had published that the atomic number of hydrogen was 1, of helium 2, etc., you could have taken that information from his writings without

infringing on his copyright. But the moment he organized the elements in a periodic table which by its organization gave deeper insight in the laws of nature, he created a copyrightable document, and everyone else reprinting that table needed his permission to do so. “Facts themselves are not copyrightable, but only the way in which information is expressed” [74]. Granted, the dividing line between facts and the expression of facts is very fuzzy and almost invisible, reason why, if you want to make your data available as open resources, you better explicitly give away the pearls and the sow's ear.

If you are willing to share your data (the “facts”, their creative arrangement, and the metadata) with others, you have to make very clear what they are allowed to do with your scientific creation, and what rights (copy rights, database rights, and whatever other rights lawyers have cooked up), if any, you preserve. If you do not issue a clear license that is legally binding on yourself and the other party, these other scientists may be afraid to touch your data for fear of being sued. Of course, every time another scientist wants to use your data the two of you could negotiate an agreement, run it by your lawyers, and sign it in the presence of witnesses. That is an expensive way of doing business, and the existence of hundreds if not thousands of unique documents would make data sharing neigh impossible in an age where a secondary analysis project such as, e.g., an IPDMA may require tens of datasets.

The solution is simple: standardized licenses, which are made part of the metadata and can be read by a computer without a law degree. Now a would-be data user can instruct her computer to go and download all datasets with a particular license (or with an even broader mandate), and she can start dealing with the scientific issue that has her interest. Such licenses are prepared by a number of organizations, but presumably the best known are those of Creative Commons. Box 2 summarizes the 6 basic license types offered by Creative Commons, indicating that releasing rights to data is not an all-or-nothing thing: the investigator still can attach certain conditions that are legally enforceable and that prohibit various form of “abuse” or misuse of the data. For research data, the CC-BY license seems the most appropriate: it places the data in the public domain (which legally does not exist in some countries, though) and enforces the academic tradition of crediting the person who did the innovative work of data creation.

Before you sign away all rights to your data and the particular format you have organized them in, it makes sense to obtain advice or even prior approval from your organization's grants/contracts office, the IRB/HSRC, the general counsel, or even the library, where copyright issues are bread and butter. It also may make sense to check with the grant maker (if any) and the publisher of papers based on the same dataset that you already have distributed, if any. If you published your data as

SDC, and gave the publisher the copyright to the article, he also owns the copyright to the data, and you cannot give these rights away without infringing on his copyright. And it may not hurt to also have a discussion with any co-investigators and other staff who made a creative contribution to the research. Giving away rights that are not yours to give is a great way to create problems all around.

## Conclusion

SCI researchers who pay attention to trends in the research world may have noticed that a culture shift is under way, with data sharing through “open data” likely becoming the “new normal” [46]. Even in frontrunner areas such as genomics [75] and especially for data from research other than clinical trials [76], a lot of changes are still needed for that (r)evolution to be complete—both in the publishing of data and the using of published data. As argued by Callahan et al. [71] and others, this includes consensus of research disciplines around terms and definitions (ontologies, CDEs), further development of data stewardship policies and data sharing infrastructure (including data security and quality control on data creation, description, and use), protection of intellectual property and appropriate citation of data, and user-friendly interfaces for data upload and download [71, 77].

In addition to these “tools”, we need models for building into our day-to-day work flow the creation of metadata and the uploading of data to repositories, and “automating” it [22], so that, whether we are preclinical lab researchers or clinician-scientists, the creation of open data is no longer an extra burden imposed from the outside, but part of good everyday science.

More carrots are also needed, including fair ways of weighting data sharing and reusing of open data in evaluating scientists for promotion in their place of work and in the scientific world as a whole, including selection criteria for science prizes. Additional sticks will help too: funders scoring data sharing plans in determining worthwhile grant proposals; university departments insisting that their faculty members actually upload the datasets they have, etc.

Lastly, and part of the culture shift that is not yet fully appreciated, is a need for more emphasis in science on validation and replication, and less idealizing (idolizing?) innovation and creativity. The latter will and should never disappear, but open data may help temper its pathologies so that more of our findings will stand up to the test of time [78]. Even those who see dangers in a de-emphasis on creativity would do good to ask themselves when starting a new research effort: “Is there any reason **not** to share the data from this research?” Sharing should become our default stance. Open data may be good for us—individually [11], as members of the scientific community, and as humans.

**Box 2** The 6 basic license types produced by Creative Commons—symbol, designation, abbreviation, and description**Attribution (CC BY)**

Lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.

**Attribution-ShareAlike (CC BY-SA)**

Lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms.

**Attribution-NoDerivs (CC BY-ND)**

Allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

**Attribution-NonCommercial (CC BY-NC)**

Lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

**Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)**

Lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

**Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)**

The most restrictive of Creative Commons' six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Dijkers M. Reduce, reuse, recycle: good stewardship of research data. *Spinal Cord*. 2019. <https://doi.org/10.1038/s41393-019-0246-8>
2. re3data. Home | re3data.org. 2018; [www.re3data.org/](http://www.re3data.org/).
3. Sudlow R, Branson J, Friede T, Morgan D, Whately-Smith C. EFSPI/PSI working group on data sharing: accessing and working with pharmaceutical clinical trial patient level datasets—a primer for academic researchers. *BMC Med Res Methodol*. 2016;16 (Suppl 1):73.
4. Martin EG, Law J, Ran W, Helbig N, Birkhead GS. Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on 3 open data platforms. *J Public Health Manag Pract*. 2017;23:e5–e13.
5. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. *J Am Med Assoc*. 2016;315:1283–4.
6. Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen-Windhouwer S, Manghi P. Navigating the unfolding open

- data landscape in ecology and evolution. *Nat Ecol Evol.* 2018;2:420–6.
7. Kobayashi S, Kane TB, Paton C. The privacy and security implications of open data in healthcare. *Yearb Med Inform.* 2018;27:41–47.
  8. Ferguson AR, Nielson JL, Cragin MH, Bandrowski AE, Martone ME. Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nat Neurosci.* 2014;17:1442–7.
  9. Morey RD, Chambers CD, Etschells PJ, et al. The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review. *R Soc Open Sci.* 2016;3:150547.
  10. Munafò MR, Nosek BA, Bishop DVM, et al. A manifesto for reproducible science. *Nat Hum Behav.* 2017;1:0021.
  11. McKiernan EC, Bourne PE, Brown CT, et al. How open science helps researchers succeed. *eLife.* 2016;5:e16800.
  12. Hesse BW. Can psychology walk the walk of open science? *Am Psychol.* 2018;73:126–37.
  13. The FAIR data principles | FORCE11. 2014. <https://www.force11.org/group/fairgroup/fairprinciples>.
  14. Giglia E, Swan A. Open access to data for a new, open science. *Eur J Phys Rehabil Med.* 2012;48:713–6.
  15. Babbage DR. Open and abundant data is the future of rehabilitation and research. *Arch Phys Med Rehabil.* 2014; 95:795–8.
  16. Data science at NIH. 2018. <https://www.ncbi.nlm.nih.gov/pubmed/>.
  17. Vickers AJ. Making raw data more widely available. *Br Med J.* 2011;342:d2323.
  18. Krumholz HM. Why data sharing should be the expected norm. *Br Med J.* 2015;350:h599.
  19. De Hert S, Samama CM. Open access to research data: the two sides of the coin. *Eur J Anaesthesiol.* 2017;34:1–3.
  20. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet.* 2014;383:166–75.
  21. Dechartres A, Ravaud P. Better prioritization to increase research value and decrease waste. *BMC Med.* 2015;13:244.
  22. Rouder JN. The what, why, and how of born-open data. *Behav Res Methods.* 2016;48:1062–9.
  23. Candela L, Castelli D, Manghi P, Tani A. Data journals: a survey. *J Assoc Inf Sci Technol.* 2015;66:1747–62.
  24. Piwowar H. Altmetrics: value all research products. *Nature.* 2013;493:159.
  25. Piwowar HA, Vision TJ. Data reuse and the open data citation advantage. *PeerJ.* 2013;1:e175.
  26. ADNI | Access Data. 2018. <http://adni.loni.usc.edu/data-samples/access-data/>.
  27. ICMJE | Home. 2018. <http://www.icmje.org/>.
  28. Haug CJ. From patient to patient—sharing the data from clinical trials. *New Eng J Med.* 2016;374:2409–11.
  29. Merson L, Gaye O, Guerin PJ. Avoiding data dumpsters—toward equitable and useful data sharing. *New Eng J Med.* 2016; 374:2414–5.
  30. Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions. *PLoS ONE.* 2011;6:e21101.
  31. Cavoukian A, El-Emam K. Dispelling the myths surrounding de-identification: anonymization remains a strong tool for protecting privacy. 2011. <https://fpf.org/wp-content/uploads/2011/07/Dispelling%20the%20Myths%20Surrounding%20De-identification%20Anonymization%20Remains%20a%20Strong%20Tool%20for%20Protecting%20Privacy.pdf>.
  32. Longo DL, Drazen JM. Data sharing. *New Eng J Med.* 2016;374:276–7.
  33. Murray MJ. Thanks for sharing: the bumpy road towards truly open data. *Br Med J.* 2016;352:i849.
  34. Charlifue S, Tate D, Biering-Sorensen F, et al. Harmonization of databases: a step for advancing the knowledge about spinal cord injury. *Arch Phys Med Rehabil.* 2016;97:1805–18.
  35. Hlatky MA, Januzzi JL Jr. Manuscripts based on datasets shared by clinical research studies. *J Am Coll Cardiol.* 2017;69:1983–5.
  36. NIH data sharing policy—Main page. 2018. [https://grants.nih.gov/grants/policy/data\\_sharing/](https://grants.nih.gov/grants/policy/data_sharing/).
  37. Dissemination and sharing of research results | NSF—National Science Foundation. 2018. <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
  38. Open access policy. 2018. <https://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy>.
  39. Wellcome Trust. Our new policy on sharing research data: What it means for you. 2018. <https://wellcome.ac.uk/news/our-new-policy-sharing-research-data-what-it-means-you>.
  40. Access to research results: Guiding principles. [http://www.science.gc.ca/eic/site/063.nsf/eng/h\\_9990CB6B.html](http://www.science.gc.ca/eic/site/063.nsf/eng/h_9990CB6B.html).
  41. OECD Principles and guidelines for access to research data from public funding. 2007. <http://www.oecd.org/science/inno/38500813.pdf>.
  42. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Ann Int Med.* 2017;167:63–65.
  43. PLOS ONE: Accelerating the publication of peer-reviewed science. <https://journals.plos.org/plosone/s/data-availability>.
  44. Research data support | Authors | Springer Nature. 2018. [https://www.springernature.com/gp/authors/research-data-policy/?utm\\_source=Other\\_digital&utm\\_medium=Other\\_digital&utm\\_content=MatAst-Nature-SNRDS\\_home-Multidisciplinary-Global&utm\\_campaign=SCIDATA\\_SUB\\_SNRDS\\_home](https://www.springernature.com/gp/authors/research-data-policy/?utm_source=Other_digital&utm_medium=Other_digital&utm_content=MatAst-Nature-SNRDS_home-Multidisciplinary-Global&utm_campaign=SCIDATA_SUB_SNRDS_home).
  45. Spinal Cord. Guide to authors. 2018. [https://images.nature.com/full/nature-assets/sc/sc\\_new\\_gta.pdf](https://images.nature.com/full/nature-assets/sc/sc_new_gta.pdf).
  46. Committee on Strategies for Responsible Sharing of Clinical Trial Data, Institute of Medicine. Washington (DC): Strategies for responsible sharing of clinical trial data. National Academies Press (US); 2015.
  47. American Psychological Association. Ethical principles of psychologists and code of conduct. 2018. <http://www.apa.org/ethics/code/index.aspx>.
  48. Naik G. Peer-review activists push psychology journals towards open data. *Nature.* 2017;543:161.
  49. Stuart D, Baynes G, Hrynaszkiewicz I, et al. Practical challenges for researchers in data sharing. 2018. [https://figshare.com/articles/Whitepaper\\_Practical\\_challenges\\_for\\_researchers\\_in\\_data\\_sharing/5975011](https://figshare.com/articles/Whitepaper_Practical_challenges_for_researchers_in_data_sharing/5975011).
  50. Alter GC, Vardigan M. Addressing global data sharing challenges. *J Empir Res Hum Res Ethics.* 2015;10:317–23.
  51. Cummings JA, Zagrodny JM, Day TE. Impact of open data policies on consent to participate in human subjects research: discrepancies between participant action and reported concerns. *PLoS ONE.* 2015;10:e0125208.
  52. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: a qualitative study. *J Med Internet Res.* 2016;18:e66.
  53. Hume S, Aerts J, Sarnikar S, Huser V. Current applications and future directions for the CDISC Operational Data Model standard: a methodological review. *J Biom Inform.* 2016;60:352–62.
  54. Biering-Sorensen F, Alai S, Anderson K, et al. Common data elements for spinal cord injury clinical research: a National Institute for Neurological Disorders and Stroke project. *Spinal Cord.* 2015;53:265–77.
  55. Dijkers MP. IPDMA: individual patient/participant data meta-analysis. *KT Update* 2016;4. [http://ktdr.org/products/update/v4n7/dijkers\\_ktupdate\\_v4n7\\_508.pdf](http://ktdr.org/products/update/v4n7/dijkers_ktupdate_v4n7_508.pdf).

56. Loder E Data sharing: making good on promises. *Br Med J*. 2018;360:k710. <https://doi.org/10.1136/bmj.k710>.
57. External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. 2017. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Regulatory\\_and\\_procedural\\_guideline/2017/09/WC500235371.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2017/09/WC500235371.pdf).
58. Dyke SO, Saulnier KM, Pastinen T, Bourque G, Joly Y. Evolving data access policy: the Canadian context. *Facets*. 2016;1:138–47.
59. Hopkins C, Sydes M, Murray G, et al. UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. *J Clin Epidemiol*. 2016;70:17–25.
60. Hrynaskiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: Guidance for journal editors, authors, and peer reviewers. *Br Med J*. 2010;340:c181.
61. Figshare. Figshare - credit for all your research. 2018. <https://figshare.com/>.
62. Dryad. Dryad digital repository. 2018. <https://www.datadryad.org/>.
63. The scholarly data archive (SDA) at Indiana University. 2018. <https://kb.iu.edu/d/aiyi>.
64. About the research data archive - University of Bath Research Data Archive. 2018. <https://researchdata.bath.ac.uk/information.html>.
65. Korea Social Science Data Archive | Seoul National University Asia Center. 2018. <http://snuac.snu.ac.kr/eng/index.php/research/korea-social-science-data-archive/>.
66. Welcome to the South African Data Archive. 2018. <http://sada.nrf.ac.za/deposit.html>.
67. Index | FITBIR: Federal Interagency Traumatic Brain Injury Research Informatics System. 2018. <https://fitbir.nih.gov/>.
68. Wicherts JM, Borsboom D, Kats J, Molenaar D. The poor availability of psychological research data for reanalysis. *Am Psychol*. 2006;61:726–8.
69. Noonan VK, Chan E, Bassett-Spiers K, et al. Facilitators and barriers to international collaboration in spinal cord injury: results from a survey of clinicians and researchers. *J Neurotrauma*. 2018;35:478–85.
70. Kramer JLK, Geisler F, Ramer L, Plunet W, Cragg JJ. Open access platforms in spinal cord injury: existing clinical trial data to predict and improve outcomes. *Neurorehabil Neural Repair*. 2017;31:399–401.
71. Callahan A, Anderson KD, Beattie MS, et al. Developing a data sharing community for spinal cord injury research. *Exp Neurol*. 2017;295:135–43.
72. DCC. Digital Curation Centre | Because good research needs good data. 2018. <http://www.dcc.ac.uk/>.
73. Burton PR, Murtagh MJ, Boyd A, et al. Data safe havens in health research and healthcare. *Bioinformatics*. 2015;31:3241–8.
74. Databases, data sets, and data accessibility—views and practices of scholarly publishers. A statement by the International Association of Scientific, Technical and Medical Publishers (STM) and the Association of Learned and Professional Society Publishers (ALPSP), 2006. [https://www.stm-assoc.org/2006\\_06\\_01\\_STM\\_ALPSP\\_Data\\_Statement.pdf](https://www.stm-assoc.org/2006_06_01_STM_ALPSP_Data_Statement.pdf).
75. Kaye J, Hawkins N. Data sharing policy design for consortia: challenges for sustainability. *Genome Med*. 2014;6:4.
76. McDonald L, Schultze A, Simpson A, Graham S, Wasiak R, Ramagopalan SV. Lack of data sharing in observational studies. *Br Med J*. 2017;359:j4866.
77. Gymrek M, Farjoun Y. Recommendations for open data science. *Gigascience*. 2016;5:22.
78. Nosek BA, Spies JR, Motyl M. Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspect Psychol Sci*. 2012;7:615–31.