




Publishers' Responsibilities in Promoting Data Quality and Reproducibility

Iain Hrynaszkiewicz 

Contents

- 1 Introduction
 - 2 Understanding Researchers' Problems and Motivations
 - 2.1 Understanding Motivations to Share Data
 - 3 Raising Awareness and Changing Behaviours
 - 3.1 Journal Policies
 - 3.2 Effectiveness of Journal Research Data Policies
 - 4 Improving the Quality, Transparency and Objectivity of the Peer-Review Process
 - 4.1 Implementation of Reporting Guidelines
 - 4.2 Editorial and Peer-Review Procedures to Support Transparency and Reproducibility
 - 4.3 Image Manipulation and Plagiarism Detection
 - 5 Better Scholarly Communication Infrastructure and Innovation
 - 5.1 Tackling Publication (Reporting) Bias
 - 5.2 Research Data Repositories
 - 5.3 Research Data Tools and Services
 - 5.4 Making Research Data Easier to Find
 - 6 Enhancing Incentives
 - 6.1 New Types of Journal and Journal Article
 - 6.2 Data and Software Citation
 - 6.3 Digital Badges for Transparency: A New Type of Incentive
 - 7 Making Research Publishing More Open and Accessible
 - 7.1 Open Access and Licencing Research for Reuse
 - 7.2 Open Publisher (Meta)Data
 - 7.3 Open for Collaboration
- References

I. Hrynaszkiewicz (✉)
Public Library of Science (PLOS), Cambridge, UK
e-mail: ihrynaszkiewicz@plos.org

Abstract

Scholarly publishers can help to increase data quality and reproducible research by promoting transparency and openness. Increasing transparency can be achieved by publishers in six key areas: (1) understanding researchers' problems and motivations, by conducting and responding to the findings of surveys; (2) raising awareness of issues and encouraging behavioural and cultural change, by introducing consistent journal policies on sharing research data, code and materials; (3) improving the quality and objectivity of the peer-review process by implementing reporting guidelines and checklists and using technology to identify misconduct; (4) improving scholarly communication infrastructure with journals that publish all scientifically sound research, promoting study registration, partnering with data repositories and providing services that improve data sharing and data curation; (5) increasing incentives for practising open research with data journals and software journals and implementing data citation and badges for transparency; and (6) making research communication more open and accessible, with open-access publishing options, permitting text and data mining and sharing publisher data and metadata and through industry and community collaboration. This chapter describes practical approaches being taken by publishers, in these six areas, their progress and effectiveness and the implications for researchers publishing their work.

Keywords

Data sharing · Open access · Open science · Peer review · Publishing · Reporting guidelines · Reproducible research · Research data · Scholarly communication

1 Introduction

Scholarly publishers have a duty to maintain the integrity of the published scholarly record. Science is often described as self-correcting, and when errors are identified in the published record, it is the responsibility of publishers to correct them. This is carried out by publishing corrections, expressions of concern or, sometimes, retracting published articles. Errors in published research can be honest, such as typographical errors in data tables or broken links to source material, but errors also result from research misconduct, including fraudulent or unethical research, and plagiarism. Only a small fraction – less than 0.1% (Grieneisen and Zhang 2012) – of published research is retracted, and papers are more likely to be retracted due to misconduct, than honest error (Fang et al. 2012).

However, the numbers of reported corrections and retractions do not account for the more pressing issue: that a large proportion of published – assumed accurate – research results are not reproducible, when reproducibility and replicability are tenets of science. Pharmaceutical companies have reported that fewer than 25% of the results reported in peer-reviewed publications could be reproduced in their labs (Prinz et al. 2011). A survey of 1,500 researchers found that more than half of respondents could not reproduce their own results and more than 70% could not reproduce the results of others (Baker 2016). An economic analysis in 2015

Table 1 Causes of poor reproducibility and poor data quality in preclinical research

	Relevant chapters elsewhere in this textbook
<i>Conduct of research</i>	
Experimental design	Chapters “Guidelines and Initiatives for Good Research Practice”, “Learning from Principles of Evidence-Based Medicine to Optimize Nonclinical Research Practices”, “General Principles of Preclinical Study Design”, “Blinding and Randomization”, “Out of Control? Managing Baseline Variability in Experimental Studies with Control Groups”, “Building Robustness Into Translational Research”, and “Design of Meta-Analysis Studies”
Quality control	Chapters “Quality of Research Tools”, “Quality Governance in Biomedical Research”, and “Costs of Implementing Quality in Research Practice”
Lab supervision and training	
Adherence to ethical standards	Chapter “Good Research Practice: Lessons from Animal Care and Use”
Culture of publishing some results, and not others	Chapter “Resolving the Tension Between Exploration and Confirmation in Preclinical Biomedical Research”
<i>Reporting of research</i>	
Completeness of methods descriptions	Chapters “Minimum Information and Quality Standards for Conducting, Reporting, and Organizing In Vitro Research”, and “Minimum Information in In Vivo Research”
Accuracy of images, figures and graphs	
Availability of research data, protocols, computer code	Chapters “Quality of Research Tools”, “Electronic Lab Notebooks and Experimental Design Assistants”, and “Data Storage”
Statistical reporting	Chapter “A Reckless Guide to P-Values: Local Evidence, Global Errors”
Publication of all scientifically sound results, regardless of their outcome	

estimated that irreproducible preclinical research costs US \$28 billion per year (Freedman et al. 2015).

There are numerous causes of irreproducibility and suboptimal data quality (Table 1). Some of these causes relate to how research is conducted and supervised, and others relate to how well or completely research is reported. Data quality and reproducibility cannot be assessed without complete, transparent reporting of research and the availability of research outputs which can be reused. Scholarly publishers have a responsibility to promote reproducible research (Hrynaszkiwicz et al. 2014) but are more able to influence the reporting of research than the conduct of research. Transparency is a precursor to reproducibility and can be supported by journals and publishers (Nature 2018).

Implementation of greater transparency in, reporting and reuse potential of, research by publishers can be achieved in several ways:

1. Understanding researchers' problems and motivations
2. Raising awareness and changing behaviours
3. Improving the quality, transparency and objectivity of the peer-review process
4. Better scholarly communication infrastructure and innovation
5. Enhancing incentives
6. Making research publishing more open and accessible

This chapter describes practical approaches being taken by publishers, in these six areas, to achieve greater transparency and discusses their progress and effectiveness and the implications for researchers.

2 Understanding Researchers' Problems and Motivations

Publishers wishing to increase transparency and reproducibility need to understand the problems (or “challenges”) researchers have in practising reproducible and transparent research. Sharing of research data is essential for reproducible research, and between 2016 and 2018, several large surveys of researchers were conducted by publishing and publishing technology companies, providing insights into researchers' reported data sharing practices and behaviours, as well as insight into what motivates researchers to share, or not share, research data.

A survey conducted in 2017, and published in 2018, by the publisher Springer Nature explored the “practical challenges” researchers have in data sharing, which received 7,719 responses, one of the largest of its kind. Seventy-six percent of respondents reported that the discoverability of their research data is important to them, and 63% had previously shared data associated with a peer-reviewed article. However, researchers also reported common problems in sharing their data, including difficulties in “organising data in a presentable and useful way” (46% of respondents), being unsure about licencing and copyright of data (37%) and not knowing which data repository to use (33%). A lack of time (26%) and being unable to cover costs of data sharing (19%) were also commonly cited (Stuart et al. 2018) (Fig. 1).

Disciplinary differences were also identified in the survey. Biological science researchers reported the highest levels of data sharing (75%), and medical science researchers reported that copyright and licencing (data ownership) issues were their biggest challenge. Medical science researchers were also most likely to report concerns about data sensitivity and misuse, and concerns about protecting research participants, consistent with other surveys (Rathi et al. 2012) of clinical researchers.

Surveys from the publishers Wiley (Wiley Open Science Researcher Survey 2016) and Elsevier (Berghmans et al. 2017) and the publishing technology company Digital Science (Science et al. 2017) have found similar results regarding the proportion of researchers who report they share data and the ways in which

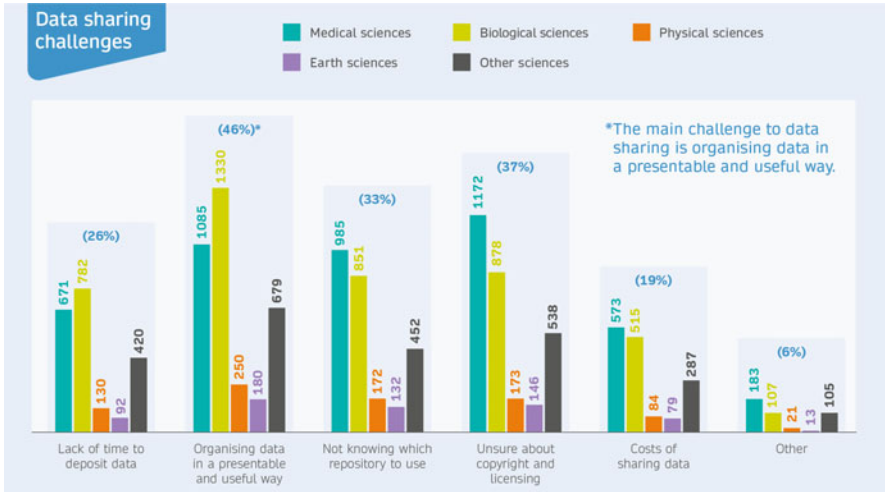


Fig. 1 Organising data in a presentable and useful way was the most common problem for researchers in data sharing, in a large survey ($n = 7,719$). Figure adapted from Stuart et al. (2018)

researchers share data (Table 2). The most common ways of sharing data that were reported tend to be suboptimal, with email being the most common method for private data sharing (Allin 2018) and journal supplementary materials being most common for public data sharing (email is not secure enough for private data sharing; data repositories are preferred over supplementary materials for public data sharing) (Michener 2015).

2.1 Understanding Motivations to Share Data

Sharing research data has been associated with an increase in the number of citations that researchers' papers receive (Piwowar et al. 2007; Piwowar and Vision 2013; Colavizza et al. 2019) and an increase in the number of papers that research projects produce (Pienta and Alter 2010). Some researchers report that increased academic credit (Science et al. 2017), and increased visibility of their research (Wiley Open Science Researcher Survey 2016; Schmidt et al. 2016), motivates them to share research data. Publishers and other service providers to researchers can help to both solve problems and increase motivations, in particular those relating to academic credit, impact and visibility (see Sect. 6).

Table 2 Seventy percent of researchers report that they share data but only 26% use data repositories when the results of five large surveys are combined

Survey responses and findings	Springer Nature global survey ^a	Springer Nature Japan survey ^b	Elsevier survey	Wiley survey	Digital Science survey
Number of respondents	7,719	1,966	1,200	4,668	2,352
Year published (year conducted)	2018 (2017)	2018 (2018)	2017 (2016)	2017 (2016)	2017 (2017)
Level of data sharing reported %	63	95 ^b	64	69	60
Use of data repositories reported %	21	25	13	41	30
Most common data sharing problem	Organising data in a presentable and useful way	Concerns about misuse of data	Data ownership	Intellectual property or confidentiality issues	Data sensitivity

^aFocused on sharing data associated with peer-reviewed publications rather than data sharing in general

^bExplicitly included private (peer-to-peer) data sharing and public data sharing

3 Raising Awareness and Changing Behaviours

Scholarly publishers and journals can help to raise awareness of issues through their wide or community-focused readership – with editorials, opinion pieces and conference and news coverage. Behavioural change can be created by changing journal and publisher policies, as researchers are motivated to comply with them when submitting papers (Schmidt et al. 2016).

3.1 Journal Policies

Journal policies and guides to authors include large amounts of information covering topics from manuscript formatting, research ethics and conflicts of interest. Many journals and publishers have, since 2015, endorsed – and are beginning to implement – the Transparency and Openness Promotion (TOP) guidelines. The TOP guidelines are a comprehensive but aspirational set of journal policies and include eight modular standards, each with three levels of increasing stringency including transparency in data, code and protocols (Nosek et al. 2014). A summary table of the requirements is available in the public domain from the Center for Open Science (Table 3). Full compliance with the TOP guidelines is

Table 3 TOP guidelines summary table^a

	Not implemented	Level I	Level II	Level III
Citation standards	Journal encourages citation of data, code and materials or says nothing	Journal describes citation of data in guidelines to authors with clear rules and examples	Article provides appropriate citation for data and materials used consistent with journal's author guidelines	Article is not published until providing appropriate citation for data and materials following journal's author guidelines
Data transparency	Journal encourages data sharing or says nothing	Article states whether data are available and, if so, where to access them	Data must be posted to a trusted repository. Exceptions must be identified at article submission	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication
Analytic methods (code) transparency	Journal encourages code sharing or says nothing	Article states whether code is available and, if so, where to access it	Code must be posted to a trusted repository. Exceptions must be identified at article submission	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication
Research materials transparency	Journal encourages materials sharing or says nothing	Article states whether materials are available and, if so, where to access them	Materials must be posted to a trusted repository. Exceptions must be identified at article submission	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing	Journal articulates design transparency standards	Journal requires adherence to design transparency standards for review and publication	Journal requires and enforces adherence to design transparency standards for review and publication
Study preregistration	Journal says nothing	Article states whether preregistration of study exists and,	Article states whether preregistration of study exists and, if so, allows journal	Journal requires preregistration of studies and provides link and badge in article to

(continued)

Table 3 (continued)

	Not implemented	Level I	Level II	Level III
		if so, where to access it	access during peer review for verification	meeting requirements
Analysis plan preregistration	Journal says nothing	Article states whether preregistration of study exists and, if so, where to access it	Article states whether preregistration with analysis plan exists and, if so, allows journal access during peer review for verification	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements
Replication	Journal discourages submission of replication studies or says nothing	Journal encourages submission of replication studies	Journal encourages submission of replication studies and conducts results blind review	Journal uses registered reports as a submission option for replication studies with peer review prior to observing the study outcomes

^aReproduced and available from the Center for Open Science, under a Creative Commons public domain CCO waiver

typically a long-term goal for journals and publishers, and implementation of the requirements is tending to happen in progressive steps, with most progress being made initially in policies for sharing of research data.

3.1.1 Standardising and Harmonising Journal Research Data Policies

While availability of research data alone does not enable reproducible research, unavailability of data (Ioannidis et al. 2009) and suboptimal data curation (Hardwicke et al. 2018) have been shown to lead to failures to reproduce results. Historically, relatively few journals have had research data policies, and, where policies have existed, they have lacked standards and consistency, which can be confusing for researchers (authors) and research support staff (Naughton and Kernohan 2016; Barbui 2016). In 2016 Springer Nature, which publishes more than 2,500 journals, began introducing standard, harmonised research data policies to its journals (Hrynaszkiewicz et al. 2017a). Similar initiatives were introduced by some of the other largest journal publishers Elsevier, Wiley and Taylor and Francis in 2017, greatly increasing the prevalence of journal data sharing policies. These large publishers have offered journals a controlled number (usually four or five) of data policy types, including a basic policy with fewer requirements compared to the more stringent policies (Table 4).

Providing several options for journal data policy is necessary because, across multiple research disciplines, some research communities and their journals are more able to introduce strong data sharing requirements than others. In parallel to these

Table 4 Summary of Springer Nature journal data policy types and examples of journals with those policy types

Policy type/ level	Policy summary	Example journal	Weblink
1	Data sharing is encouraged	<i>Cardiovascular Drugs and Therapy</i>	https://www.springer.com/medicine/cardiology/journal/10557
2	Data sharing and evidence of data sharing and data availability statements are encouraged	<i>Clinical Drug Investigation</i>	https://www.springer.com/adis/journal/40261
3	Data sharing encouraged and data statements are required	<i>Nature</i>	http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf
4	Data sharing, evidence of data sharing, data availability statements and peer review of data required	<i>Scientific Data</i>	https://www.nature.com/sdata/policies/data-policies

individual publisher's data policy initiatives, a global collaboration of publishers, and other stakeholders in research, have created a master research data policy framework that supports all journal and publisher requirements (Hrynaskiewicz et al. 2017b, 2019).

There have also been research data policy initiatives from communities of journals and journal editors. In 2010 journals in ecology and evolutionary biology joined in supporting a Joint Data Archiving Policy (JDAP) (Whitlock et al. 2010), Public Library of Science (PLOS) introduced a strong data sharing policy to all its journals in 2014, and in 2017 the International Committee of Medical Journal Editors (ICMJE) introduced a standardised data sharing policy (Taichman et al. 2017) for its member journals, which include *BMJ*, *Lancet*, *JAMA* and the *New England Journal of Medicine*. The main requirement of the ICMJE policy was not to mandate data sharing but for reports of clinical trials to include a data sharing statement.

Data sharing statements (also known as data availability statements) are a common feature of journal and publisher data policies. They provide a statement about where data supporting the results reported in a published article can be found – including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. Many journals and publishers provide guidance on preparing data availability statements (e.g. <https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880>). All Public Library of Science (PLOS), Nature and BMC journals require data availability statements (Colavizza et al. 2019). Some research funding agencies – including the seven UK research councils (UK Research and Innovation 2011) – also require the provision of data availability statements in published articles.

Experimental pharmacology researchers publishing their work in 2019 and beyond, regardless of their target journal(s), should be prepared at minimum to provide a statement regarding the availability and accessibility of the research data that support the results of their papers.

Code and Materials Sharing Policies

To assess data quality and enable reproducibility, transparency and sharing of computer code and software (and supporting documentation) are also important – as is, where applicable, the sharing of research materials. Materials include samples, cell lines and antibodies. Journal and publisher policies on sharing code, software and materials are becoming more common but are generally less well evolved and less widely established compared to research data policies.

In 2015 the Nature journals introduced a policy across all its research titles that encourages all authors to share their code and provide a “code availability” statement in their papers (Nature 2015). *Nature Neuroscience* has taken this policy further, by piloting peer review of code associated with research articles in the journal (Nature 2017). Software-focused journals such as the *Journal of Open Research Software* and *Source Code for Biology and Medicine* tend to have the most stringent requirements for availability and usability of code.

3.2 Effectiveness of Journal Research Data Policies

Journal submission guidelines can increase transparent research practices by authors (Giofrè et al. 2017; Nuijten et al. 2017). Higher journal impact factors have been associated with stronger data sharing policies (Vasilevsky et al. 2017). Stronger data policies that mandate and verify data sharing by authors, and require data availability statements, are more effective at ensuring data are available long term (Vasilevsky et al. 2017) compared to policies that passively encourage data sharing (Vines et al. 2013). Many journal policies ask authors to make supporting data available “on reasonable request”, as a minimum requirement. This approach to data sharing may be a necessity in medical research, to protect participant privacy, but contacting authors of papers to obtain copies of datasets is an unreliable method of sharing data (Vanpaemel et al. 2015; Wicherts et al. 2006; Savage and Vickers 2009; Rowhani-Farid and Barnett 2016). Using more formal, data sharing (data use) agreements can improve authors’ willingness to share data on request (Polanin and Terzian 2018), and guidelines on depositing clinical data in controlled-access repositories have been defined by editors and publishers, as a practical alternative to public data sharing (Hrynaszkiewicz et al. 2016). Publishers are also supporting editors to improve policy effectiveness and consistency of implementation (Graf 2018).

4 Improving the Quality, Transparency and Objectivity of the Peer-Review Process

The reporting of research methods, interventions, statistics and data on harms of drugs, in healthcare research, and the presentation of results in journal articles, has repeatedly been found to be inadequate (Simera et al. 2010). Increasing the consistency and detail of reporting key information in research papers, with reporting guidelines and checklists, supports more objective assessment of papers in the peer-review process.

4.1 Implementation of Reporting Guidelines

The prevalence and endorsement of reporting guidelines, catalogued by the EQUATOR Network (<http://www.equator-network.org>), in journals has increased substantially in the last decade. Reporting guidelines usually comprise a checklist of key information that should be included in manuscripts, to enable the research to be understood and the quality of the research to be assessed. Reporting guidelines are available for a wide array of study designs, such as randomised trials (the CONSORT guideline), systematic reviews (the PRISMA guidelines) and animal preclinical studies (the ARRIVE guidelines; discussed in detail in another chapter in this volume).

The positive impact of endorsement of reporting guidelines by journals has however been limited (Percie du Sert et al. 2018), in part due to reporting guidelines often being implemented by passive endorsement on journal websites (including them in information for authors). Some journals, such as the medical journals *BMJ* and *PLOS Medicine*, have mandated the provision of certain completed reporting guidelines, such as CONSORT, as a condition of submitting manuscripts. Endorsement and implementation of reporting guidelines has been more prevalent in journals with higher impact factors (Shamseer et al. 2016). More active interventions to enforce policy in the editorial process are generally more effective, as demonstrated with data sharing policies (Vines et al. 2013), but these interventions are also more costly as they increase demands on authors' and editors' time. For larger, multidisciplinary journals publishing many types of research, identifying and enforcing the growing number of relevant reporting guidelines, which can vary from paper to paper that is submitted, is inherently more complex and time-consuming. These processes of checking manuscripts for adherence to guidelines can however be supported with artificial intelligence tools such as <https://www.penelope.ai/>.

An alternative approach to this problem taken by the multidisciplinary science journal *Nature* was to introduce a standardised editorial checklist to promote transparent reporting that could be applied to many different study designs and research disciplines. The checklist was developed by the journal in collaboration with researchers and funding agencies (Anon 2013) and is implemented by professional editors, who require that all authors complete it. The checklist elements focus on experimental and analytical design elements that are crucial for the interpretation of

research results. This includes description of methodological parameters that can introduce bias or influence robustness and characterisation of reagents that may be subject to biological variability, such as cell lines and antibodies. The checklist has led to improved reporting of risks of bias in in vivo research and improved reporting of randomisation, blinding, exclusions and sample size calculations; however in vitro data compliance was not improved, in an independent assessment of the checklist's effectiveness (Macleod and The NPQIP Collaborative Group 2017; The NPQIP Collaborative Group 2019).

In 2017 the *Nature* checklist evolved into two documents, a “reporting summary” that focuses on experimental design, reagents and analysis and an “editorial policy checklist” that covers issues such as data and code availability and research ethics. The reporting summary document is published alongside the associated paper and, to enable reuse by other journals and institutions, is made available under an open-access licence (Announcement 2017) (Fig. 2).

4.2 Editorial and Peer-Review Procedures to Support Transparency and Reproducibility

Peer review is important for assessing and improving the quality of published research (even if evidence of its effectiveness is often questioned (Smith 2010)). Most journals have been, understandably, reluctant to give additional mandatory tasks to peer reviewers – who may already be overburdened with the continuing increase in volume of publications (Kovanis et al. 2016) – to ensure journal policy compliance and assessment of research data and code. Journal policies often encourage reviewers to consider authors' compliance with data sharing policies, but formal peer review of data tends to occur only in a small number of specialist journals, such as data journals (see later in this chapter) and journals with the strictest data sharing policies. The most stringent research data policy of Springer Nature's four types of policy requires peer reviewers to access the supporting data for every publication in a journal and includes guidelines for peer reviewers of data:

Peer reviewers should consider a manuscript's Data availability statement (DAS), where applicable. They should consider if the authors have complied with the journal's policy on the availability of research data, and whether reasonable effort has been made to make the data that support the findings of the study available for replication or reuse by other researchers.

For the Data availability statement, reviewers should consider:

- Has an appropriate DAS been provided?
- Is it clear how a reader can access the data?
- Where links are provided in the DAS, are they working/valid?
- Where data access is restricted, are the access controls warranted and appropriate?
- Where data are described as being included with the manuscript and/or supplementary information files, is this accurate?

For the data files, where available, reviewers should consider:

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission, please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a Confirmed
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
 - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
 - The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
 - A description of all covariates tested
 - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
 - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
 - For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
 - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
 - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
 - Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- Data collection**
- Data analysis**

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

Fig. 2 The Nature Research Reporting Summary Checklist, available under a Creative Commons attribution licence from <https://www.nature.com/authors/policies/ReportingSummary.pdf>

- Are the data in the most appropriate repository?
- Were the data produced in a rigorous and methodologically sound manner?
- Are data and any metadata consistent with file format and reporting standards of the research community?
- Are the data files deposited by the authors complete and do they match the descriptions in the manuscript?
- Do they contain personally identifiable, sensitive or inappropriate information?

However, as of 2019 fewer than ten journals have implemented this policy of formal data peer review as a mandatory requirement, and journal policies on data sharing and reproducibility tend to focus on transparent reporting, such as including links to data sources. This enables a motivated peer reviewer to assess aspects of a study, such as data and code, more deeply, but this is not routinely expected.

In specific disciplines, journals and study designs, additional editorial assessment and statistical review are routinely employed. Some medical journals, such as *The Lancet*, consistently invite statistical review of clinical trials, and statistical reviewers have been found to increase the quality of reporting of biomedical articles (Cobo et al. 2007). Not all journals, such as those without full-time editorial staff, have sufficient resources to statistically review all research papers. Instead, journals may rely on editors and nonstatistical peer reviewers identifying if statistical review is warranted and inviting statistical review case-by-case.

Some journals have taken procedures on assessing reproducibility and transparency even further. The journal *Biostatistics* employs an Associate Editor for reproducibility, who awards articles “kite marks” for reproducibility, which are determined by the availability of code and data and if the Associate Editor for reproducibility is able to reproduce the results in the paper (Peng 2009). Another journal, *npj Breast Cancer*, has involved an additional editor, a Research Data Editor (a professional data curator), to assess every accepted article and give authors editorial support to describe and share link to the datasets that support their articles (Kirk and Norton 2019).

4.3 Image Manipulation and Plagiarism Detection

Plagiarism and self-plagiarism are common forms of misconduct and common reasons for papers being retracted (Fang et al. 2012). In the last decade, many publishers have adopted plagiarism detection software, and some apply this systematically to all submissions. Plagiarism detection software, such as iThenticate, works by comparing manuscripts against a database of billions of web pages and 155 million content items, including 49 million works from 800 scholarly publishers that participate in CrossRef Similarity Check (<https://www.crossref.org/services/similarity-check/>). Plagiarism detection is an important mechanism for publishers, editors and peer reviewers to maintain quality and integrity in the scholarly record. Although less systematically utilised in the editorial process, software for automated

detection of image manipulation – a factor in about 40% of retractions in the biomedical literature and thought to affect 6% of published papers – is also available to journals (Bucci 2018).

5 Better Scholarly Communication Infrastructure and Innovation

Publishers provide and utilise scholarly communication infrastructure, which can be both an enabler and a barrier to reproducibility. In this chapter, “scholarly communication infrastructure” means journals, article types, data repositories, publication platforms and websites, content production and delivery systems and manuscript submission and peer-review systems.

5.1 Tackling Publication (Reporting) Bias

Publication bias, also known as reporting bias, is the phenomenon in which only some of the results of research are published and therefore made available to inform evidence-based decision-making. Papers that report “positive results”, such as positive effects of drugs on a condition or disease, are more likely to be published, are more likely to be published quickly and are likely to be viewed more favourably by peer reviewers (McGauran et al. 2010; Emerson et al. 2010). In healthcare-related research, this is a pernicious problem, and widely used healthcare interventions, such as the antidepressant reboxetine (Eyding et al. 2010), have been found to be ineffective or potentially harmful, when unpublished results and data are obtained and combined with published results in meta-analyses.

5.1.1 Journals

Providing a sufficient range of journals is a means to tackle publication bias. Some journals have dedicated themselves exclusively to the publication of “negative” results, although have remained niche publications and many have been discontinued (Teixeira da Silva 2015). But there are many journals that encourage publication of all methodologically sound research, regardless of the outcome. The BioMed Central (BMC) journals launched in 2000 with this mission to assess scientific accuracy rather than impact or importance and to promote publication of negative results and single experiments (Butler 2000). Many more “sounds science” journals – often multidisciplinary “mega journals” including *PLOS One*, *Scientific Reports* and *PeerJ* – have since emerged, almost entirely based on an online-only open-access publishing model (Björk 2015). There is no shortage of journals to publish scientifically sound research, yet publication bias persists. More than half of clinical trial results remain unpublished (Goldacre et al. 2018).

5.1.2 Preregistration of Research

Preregistration of studies and study protocols, in dedicated databases, before data are collected or patients recruited, is another means to reduce publication bias. Registration is well established – and mandatory – for clinical trials, using databases such as [ClinicalTrials.gov](https://www.clinicaltrials.gov) and the ISRCTN register. The prospective registration of clinical trials helps ensure the data analysis plans, participant inclusion and exclusion criteria and other details of a planned study are publicly available before publication of results. Where this information is already in the public domain, it reduces the potential for outcome switching or other sources of bias to occur in the reported results of the study (Chan et al. 2017). Clinical trial registration has been common since 2005, when the ICMJE introduced a requirement for prospective registration of trials as a condition for publication in its member journals. Publishers and editors have been important in implementing this requirement to journals.

Preregistration has been adopted by other areas of research, and databases are now available for preregistration of systematic reviews (in the PROSPERO database) and for all other types of research, with the Open Science Framework (OSF) and the Registry for International Development Impact Evaluations (RIDIE).

Registered Reports

A more recent development for preregistration is a new type of research article, known as a registered report. Registered reports are a type of journal article where the research methods and analysis plans are both pre-registered and submitted to a journal for peer review before the results are known (Robertson 2017). Extraordinary results can make referees less critical of experiments, and with registered reports, studies can be given in principle acceptance decisions by journals before the results are known, avoiding unconscious biases that may occur in the traditional peer-review process. The first stage of the peer-review process used for registered reports assesses a study's hypothesis, methods and design, and the second stage considers how well experiments followed the protocol and if the conclusions are justified by the data (Nosek and Lakens 2014). Since 2017, registered reports began to be accepted by a number of journals from multiple publishers including Springer Nature, Elsevier, PLOS and the BMJ Group.

Protocol Publication and Preprint Sharing

Predating registered reports, in clinical trials in particular, it has been common since the mid-2000s for researchers to publish their full study protocols as peer-reviewed articles in journals such as *Trials* (Li et al. 2016). Another form of early sharing of research results, before peer review has taken place or they are submitted to a journal, is preprint sharing. Sharing of preprints has been common in physical sciences for a quarter of century or more, but since the 2010s preprint servers for biosciences ([biorxiv.org](https://www.biorxiv.org)), and other disciplines, have emerged and are growing rapidly (Lin 2018). Journals and publishers are, increasingly, encouraging their use (Luther 2017).

5.2 Research Data Repositories

There are more than 2000 data repositories listed in re3data (<https://www.re3data.org/>), the registry of research data repositories (and more than 1,100 databases in the curated FAIRsharing resource on data standards, policies and databases <https://fairsharing.org/>). Publishers' research data policies generally preference the use of third party or research community data repositories, rather than journals hosting raw data themselves. Publishers can help enable repositories to be used, more visible, and valued, in scholarly communication. This is beneficial for researchers and publishers, as connecting research papers with their underlying data has been associated with increased citations to papers (Dorch et al. 2015; Colavizza et al. 2019).

Publishers often provide lists of recommended or trusted data repositories in their data policies, to guide researchers to appropriate repositories (Callaghan et al. 2014) as well as linking to freely available repository selection tools (such as <https://repositoryfinder.datacite.org/>). Some publishers – such as Springer Nature via its Research Data Support helpdesk (Astell et al. 2018) – offer free advice to researchers to find appropriate repositories. The journal *Scientific Data* has defined criteria for trusted data repositories in creating and managing its list of data repositories and makes its recommended repository list available for reuse (Scientific Data 2019).

Where they are available, publishers generally promote the use of community data repositories – discipline-specific, data repositories and databases that are focused on a particular type or format of data such as GenBank for genetic sequence data. However, much research data – sometimes called the “long tail” of research data (Ferguson et al. 2014) – do not have common databases, and, for these data, general-purpose repositories such as figshare, Dryad, Dataverse and Zenodo are important to enable all research data to be shared permanently and persistently.

Publishers, and the content submission and publication platforms they use, can be integrated with research data repositories – in particular these general repositories – to promote sharing of research data that support publications. The Dryad repository is integrated to varying extents with a variety of common manuscript submission systems such as Editorial Manager and Scholar One. Integration with repositories makes it easier and more efficient for authors to share data supporting their papers. The journal *Scientific Data* enables authors to deposit data into figshare seamlessly during its submission process, resulting in more than a third of authors depositing data in figshare (data available via <http://scientificdata.isa-explorer.org>). Many publishers have invested in technology to automatically deposit small datasets shared as supplementary information files with journals articles into figshare to increase their accessibility and potential for reuse.

5.3 Research Data Tools and Services

Publishers are diversifying the products and services they provide to support researchers practice reproducible research (Inchcoombe 2017). The largest scholarly publisher Elsevier (RELX Group), for example, has acquired software used by

researchers before they submit work to journals, such as the electronic lab notebook Hivebench. Better connecting scholarly communication infrastructure with researchers' workflow and research tools is recognised by publishers as a way to promote transparency and reproducibility, and publishers are increasingly working more closely with research workflow tools (Hrynaszkiewicz et al. 2014).

While “organising data in a presentable and useful way” is a key barrier to data sharing (Stuart et al. 2018), data curation as a distinct profession, skill or activity has tended to be an undervalued and under-resourced in scholarly research (Leonelli 2016). Springer Nature, in 2018, launched a Research Data Support service (<https://www.springernature.com/gp/authors/research-data/research-data-support>) that provides data deposition and curation support for researchers who need assistance from professional editors in sharing data supporting their publications. Use of this service has been associated with increased metadata quality (Grant et al. 2019; Smith et al. 2018). Publishers, including Springer Nature and Elsevier, provide academic training courses in research data management for research institutions. Some data repositories, such as Dryad, offer metadata curation, and researchers can also often access training and support from their institutions and other third parties such as the Digital Curation Centre.

5.4 Making Research Data Easier to Find

Publishing platforms can promote reproducibility and provenance tracking by improving the connections between research papers and data and materials in repositories. Ensuring links between journal articles and datasets are present, functional and accurate is technologically simple, but can be procedurally challenging to implement when multiple databases are involved. Connecting published articles and research data in a standardised manner across multiple publishing platforms and data repositories, in a dynamic and universally adoptable manner, is highly desirable. This is the aim of a collaborative project between publishers and other scholarly infrastructure providers such as CrossRef, DataCite and OpenAIRE. This Scholarly Link Exchange (or, “Scholix”) project enables information on links between articles and data to be shared between all publishers and repositories in a unified manner (Burton et al. 2017). This approach, which publishers are important implementers of, means readers accessing articles on a publisher platform or literature database or data repository, such as Science Direct or EU PubMed Central or Dryad, will be provided with contemporaneous and dynamic information on datasets that are linked to articles in other journals or databases and vice versa.

6 Enhancing Incentives

Publications in peer-reviewed journals, and citations, are established mechanisms for assigning credit for scholarly contributions and for researchers and institutions to provide evidence for their research outputs and impact. Publishers can offer

Table 5 Examples of data, software, methods and protocol journals

Type of journal	Journal	Publisher
Data journal	<i>Scientific Data</i>	Springer Nature
Data journal	<i>Earth Systems Science Data</i>	Copernicus
Data journal	<i>Data in Brief</i>	Elsevier
Data journal	<i>GigaScience</i>	Oxford University Press/BGI
Software journal	<i>Journal of Open Research Software</i>	Ubiquity Press
Software journal	<i>Source Code for Biology and Medicine</i>	Springer Nature
Software journal	<i>SoftwareX</i>	Elsevier
Protocol journal	<i>Nature Protocols</i>	Springer Nature
Protocol journal	<i>Current Protocols</i>	Wiley
Methods journal	<i>Nature Methods</i>	Springer Nature
Methods journal	<i>MethodsX</i>	Elsevier

incentives to promote transparency by providing opportunities for additional articles and citations and new forms of incentive such as digital badges.

6.1 New Types of Journal and Journal Article

In the last 10 years, more journals, and types of journal article, have emerged that publish articles that describe specific parts of a research project. The print-biased format of traditional research articles does not always provide sufficient space to communicate all aspects of a research project. These new publications include journals that specialise in publishing articles that describe datasets or software (code), methods or protocols. Established journals have also introduced new article types that describe data, software, methods or protocols (Table 5).

Of these journals and article types, data journals and data papers are the most common. Data papers do not include a Results or Conclusion, like traditional research papers. They generally describe a publicly available dataset in sufficient detail so that another researcher can find, understand and reuse the data. Data journals generally do not publish raw data, but publish peer-reviewed papers that describe datasets (Hrynaszkiewicz and Shintani 2014). Data papers often include more detailed or technical information that may be excluded from traditional research papers, or which might only appear as supplementary files in traditional research papers. Data papers can both accompany traditional research papers and be independent articles that enable the publication of important datasets and databases that would not be considered as a traditional publication.

Papers published in data journals attract citations. While the number of articles published in data journals is steadily growing, they, however, represent a small proportion of the published literature overall (Berghmans et al. 2017).

6.2 Data and Software Citation

Research data, software and other research outputs, when published in digital repositories, can be assigned Digital Object Identifiers (DOIs), like research papers and chapters, enabling these research outputs to be individually discovered and cited and their citations measured in the same way.

Citing data and software promotes reproducibility by enabling linking and provenance tracking of research outputs. Papers can be persistently linked to the version (s) of data and code that were used or generated by the experiments they describe. Data citation can also provide more specific evidence for claims in papers, when those claims are based on published data. Citation of data and software is encouraged, and in some case required, as part of many journals' data sharing and reproducible research policies (Hrynaskiewicz et al. 2017a). Some funding agencies, such as the National Science Foundation in the USA, encourage researcher to list datasets and software (in addition to traditional publications) as part of their bibliographic sketches (Piwowar 2013).

From the researcher's (author's) perspective, citing data and software in reference lists is the same as citing journal articles and book chapters are cited. Several datasets are cited in this chapter, such as Smith et al. (Smith et al. 2018), and software can, similarly, also be cited when it is deposited in repositories that assign DOIs. Zenodo and figshare are commonly recommended for depositing code and software so that they can be cited.

To promote data citation and to enable data citations and links to be more visible to readers, publishers have implemented changes to the structure of published content (the XML underlying the digital version of journal articles) (Cousijn et al. 2017, 2018). Publishers and other scholarly infrastructure providers, such as DataCite and CrossRef (member organisations that generate DOIs for digital research outputs), are collaborating to enable data citation to be implemented and practised consistently, regardless of where researchers publish. Data citations, in article reference lists (bibliographies), have historically appeared in a small proportion of the published literature, but data citations have been increasing year-on-year (Garza and Fenner 2018). Researchers have indicated that they value the credit they receive through data citations, in some cases equally to the credit they receive from citations to their papers (Science et al. 2017).

6.3 Digital Badges for Transparency: A New Type of Incentive

The Center for Open Science offers digital badges that are displayed on published articles to highlight, or reward, papers where the data and materials are openly available and for studies that are pre-registered. Badges signal to the reader that the content has been made available and certify its accessibility in a persistent location. More than 40 journals, in 2018, offered or were experimenting with the award of badges to promote transparency (Blohowiak 2013). The use of digital

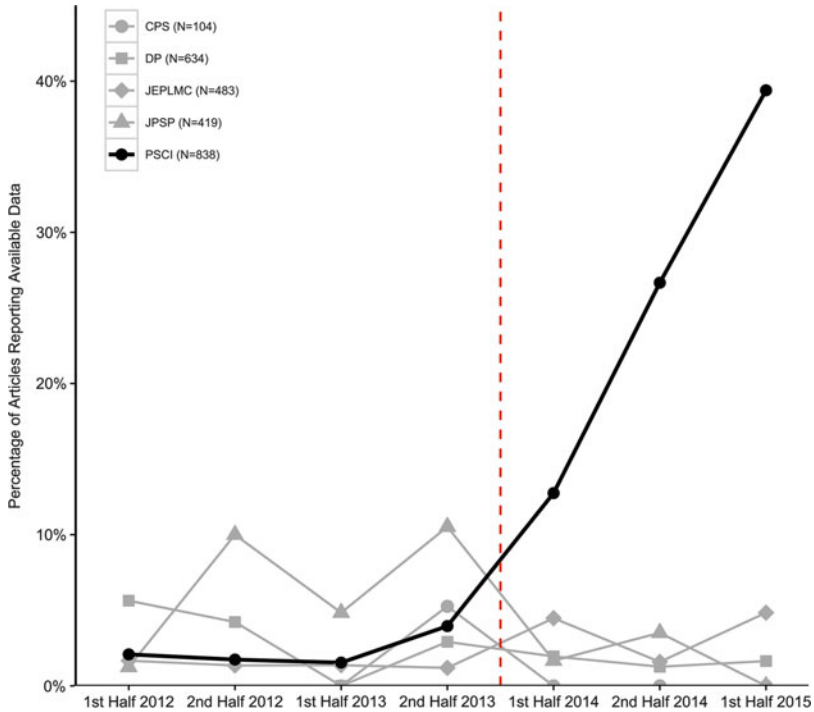


Fig. 3 Percentage of articles reporting open data by half year by journal. Darker line indicates *Psychological Science*, and dotted red line indicates when badges were introduced in *Psychological Science* and none of the comparison journals. Figure and legend reproduced from Kidwell et al. (2016)

badges is most prolific in psychology and human behavioural research journals, but they are also used in some microbiology, primatology and geoscience journals.

Awarding digital badges to authors has been associated with increased rates of data sharing by authors. When the journal *Psychological Science* (PSCI) introduced badges for articles with open data, the proportion of articles with open data increased (Fig. 3), compared to previous levels of data sharing in the journal. Data sharing also increased in the journal compared to other psychology journals (Munafò et al. 2017; Kidwell et al. 2016).

Digital badges being an effective incentive for data sharing has also been confirmed in a systematic review (Rowhani-Farid et al. 2017). The badges that are awarded by the journal *Biostatistics*' Associate Editor for reproducibility have also been associated with increased data sharing, although, in the same study, badges did not have an impact on the sharing of code (Rowhani-Farid and Barnett 2018).

Badges are usually awarded by authors self-disclosing information or they are awarded as part of the peer-review process. Another method of awarding badges adopted by *BMC Microbiology* involves the data availability statements of each paper being assessed, independently, by the publisher (Springer Nature 2018).

Box 1 Practical Recommendations for Researchers to Support the Publication of Reproducible Research

Before You Carry Out Your Research

- Check if your institution or employer, funding agency and target journals have policies on sharing and managing research data, materials and code or more broadly on reproducibility and open science.
 - Seek advice on compliance with these policies, and support including formal training, where needed.
 - Note that journal policies on data sharing are generally agnostic of whether research is industry or academically sponsored.
- Consider how you will store and manage your data and other research outputs and plan accordingly, including whether additional or specific funding is required to cover associated costs.
 - Preparing a Data Management Plan (DMP) is recommended and is often required under funding agency and institutional policies. Free tools such as <https://dmponline.dcc.ac.uk/> can assist in creating DMPs.
- Determine if there are standards and expectations, and existing infrastructure such as data repositories, for sharing data in your discipline.
 - Use resources such as <https://fairsharing.org/> to explore standards, policies, databases and reporting guidelines.
 - Establish if there are existing repositories for the type of data you generate. Where they exist use discipline-specific repositories for your data, and general repositories for other data types.
- Familiarise yourself with tools that enable reproducibility, and version control, particularly for computational work (Markowetz 2015).
- Where appropriate databases exist, consider preregistration of your study (for clinical trials registration in a compliant database is mandatory) as a means to reduce the potential for bias in analyses.
- For clinical studies in particular, publish your study protocol as a peer-reviewed article, or at minimum be prepared to share it with journal editors and peer reviewers.
 - Free tools such as <https://www.nature.com/protocolexchange/> and <https://www.protocols.io/> can be used to share methodological knowledge.
- If your target journal(s) offer them, consider preparing a registered report.

When Preparing to Submit Your Research Results to a Journal

- Register for an ORCID identifier and encourage your co-authors to do the same.
- Publish a preprint of your paper in a repository such as bioRxiv, enabling the community to give you feedback on your work and for you to assert ownership and claim credit for your work early.

(continued)

Box 1 (continued)

- Prepare your data and code for deposition in a repository, and make these available to editors and peer reviewers.
 - Use repositories rather than supplementary information files for your datasets and code.
- Consider publishing data papers, software papers or methods-focused papers to complement your traditional research papers, particularly if detailed information that enables understanding and reuse your research does not form part of your traditional papers.
- If the results of your research are inconclusive and show no difference between comparison groups (“negative results”), publish them. Many journals consider such papers.
- Always include clear statements in your publications about the availability of research data and code generated or utilised by your research.
 - If there are legitimate restrictions on the availability and reuse of your data, explain them in your data availability statements.
 - Wherever possible, include links to supporting datasets in your publications – this supports reproducibility and is associated with increased citations to papers.
- Be prepared to share with editors and peer reviewers any materials supporting your papers that might be needed to verify, replicate or reproduce the results.
 - Many repositories enable data to be shared privately before publication and in a way that protects peer reviewers’ anonymity (where required).
- Cite, in your reference lists and bibliographies, any persistent, publicly available datasets that were generated or reused by your research.

After Publication of Your Research

- Be prepared to respond to reasonable requests from other scientists to reuse your data.
 - Non-compliance with data sharing policies of journals can lead to corrections, expressions of concern or retractions of papers.
- Try to view the identification of honest errors in published work – yours and others – as a positive part of the self-correcting nature of science.
- Remember working transparently and reproducibly is beneficial to your own reputation, productivity and impact as a researcher, as well as being beneficial to science and society (Markowetz 2015).

7 Making Research Publishing More Open and Accessible

The sixth and final area in which publisher can promote transparency relates to how open and accessible publishers are as organisations. This refers firstly to the content publishers distribute and secondly to the accessibility of other information and resources from publishers.

7.1 Open Access and Licencing Research for Reuse

Publishing more research open access, so that papers are freely and immediately available online, is an obvious means to increase transparency. The proportion of the scholarly literature that is published open access, each year, is increasing by 10–15%. Open access accounted for 17% of published articles in 2015 (Johnson et al. 2017), and the two largest journals in the world – *Scientific Reports* and *PLOS One* – are open-access journals.

Open-access publishing means more than access to research; it is also about promoting free reuse and redistribution of research, through permissive copyright licences (Suber 2012). Open-access journals and articles are typically published under Creative Commons attribution licences, such as CC BY, which means that the work can be copied, distributed, modified and adapted freely, by anyone, provided the original authors are attributed (the figures in this chapter are examples of this practice).

Publishing research under CC BY, or equivalent copyright licences, is important for promoting reproducibility of research because it enables published research outputs to be reused efficiently, by humans and machines. With this approach, the pace of research need not be slowed by the need to negotiate reuse rights and agreements with researchers and institutions. Meanwhile scholarly norms of acknowledging previous work (through citation) and legal requirements for attribution in copyright will ensure that researchers are credited for their contributions (Hrynaszkiewicz and Cockerill 2012).

Reuse of the research literature is essential for text and data mining research, and this kind of research can progress more efficiently with unrestricted access to and reuse of the published literature. Publishers can enable the reuse of research content published in subscription and open-access journals with text and data mining policies and agreements. Publishers typically permit academic researchers to programmatically access their publications, such as through secure content application programming interfaces (APIs), for text and data mining research (Text and Data Mining – Springer; Text and Data Mining Policy – Elsevier).

7.2 Open Publisher (Meta)Data

For other kinds of content, including research data, publishers can promote ease of access and reuse by applying and setting standards for content licences that enable

reuse easily. In 2006 multiple publishers signed a joint statement agreeing not to take copyright in research data (STM, ALPSP 2006). Publishers have also promoted the use of liberal, public domain legal tools for research data and metadata. The publisher BMC introduced, in 2013, a default policy whereby any data published in their more than 250 journals would be available in the public domain, under the Creative Commons CC0 waiver (Hrynaszkiewicz et al. 2013). Publishers can also make data about their content catalogues (metadata) openly available. Springer Nature's SciGraph, for example, is a linked open data platform for the scholarly domain and collates metadata from funders, research projects, conferences, affiliations and publications (SciGraph). Many publishers also make the bibliographies (reference lists) of all their publications, subscription and open access, available openly as "open citations" (Shotton 2013; I4OC).

Beyond published articles, journals and associated metadata, publishers can share other information openly. This includes survey findings (Table 1) and the results of projects to improve transparency reproducibility – such as around data sharing policies (Hrynaszkiewicz et al. 2017a) and research data curation (Smith et al. 2018). Resources produced and curated by publishers can also be made available to the wider community (such as Scientific Data 2019).

7.3 Open for Collaboration

Publishers can promote transparency through collaboration. The biggest policy and infrastructural challenges that enable the publication of more reproducible research can only be tackled by multiple publishers collaborating as an industry and collaboration with other organisations that support the conduct and communication of research – repositories, institutions and persistent identifier providers. Progress resulting from such collaborations has been seen in data citation (Cousijn et al. 2017), data policy standardisation (Hrynaszkiewicz et al. 2017b), reporting standards to enhance reproducibility (McNutt 2014) and provenance tracking of research outputs and researchers, through persistent identification initiatives such as ORCID (<https://orcid.org/organizations/publishers/best-practices>). All of which, combined, help publishers and the wider research community to make practical improvements to the communication of research that support improved data quality and reproducibility.

7.3.1 The Future of Scholarly Communication?

In some respects the future of scholarly communication is already here, with dynamic, reproducible papers (Lewis et al. 2018), workflow publication, data integration and interactive data, figures and code all possible, albeit at a relatively small scale. However, these innovations remain highly unevenly distributed, and the majority of published scholarly articles remain largely static objects, with the PDF format remaining popular with many readers. Like most scientific advances, progress in scholarly communication tends not to be made through giant leaps of progress

but by slow, steady, incremental improvements. However, numerous major publishers have expressed strong support for open science and are introducing practical measures to introduce and strengthen policies on transparency of all research outputs, as a prerequisite to improving reproducibility. Researchers should expect continued growth in transparency policies of journals and be prepared for demands for more transparency in the reporting of their research (see Box 1 for practical suggestions for researchers). Increasing computerisation and machine readability of papers, with integration of data and code and enhancement of metadata increasing, will promote reproducibility and new forms of research quality assessment. This will help the research community assess individual research projects more specifically than the inappropriate journal-based measure of the impact factor. Large publishers will continue to diversify the types of content they publish and diversify their businesses, evolving into service providers for researchers and institutions and including content discovery, research metrics, research tools, training and analytics in their activities alongside publishing services. Technology and services are just part of implementing reproducible research, and cultural and behavioural change – and demonstrating value and impact of reproducible research – will continue to be incentivised with policies of all stakeholders in research. Monitoring compliance with transparency and reproducibility policies remains a challenge, but increasing standardisation of policies will enable economies of scale in monitoring compliance.

Competing Interests At the time of writing this chapter, the author (IH) was employed by Springer Nature. Since July 2019 the author is employed by Public Library of Science (PLOS). Neither employer had any role in the preparation or approval of the chapter.

References

- Allin K (2018) Research data: challenges and opportunities for Japanese researchers – Springer Nature survey data. https://figshare.com/articles/Research_data_challenges_and_opportunities_for_Japanese_researchers_-_Springer_Nature_survey_data/6328952/1
- Announcement (2017) Towards greater reproducibility for life-sciences research in Nature. *Nature* 546:8
- Anon J (2013) Announcement: reducing our irreproducibility. *Nature* 496:398–398
- Astell M, Hrynaszkiewicz I, Grant R, Smith G, Salter J (2018) Have questions about research data? Ask the Springer Nature helpdesk. https://figshare.com/articles/Providing_advice_and_guidance_on_research_data_a_look_at_the_Springer_Nature_Helpdesk/5890432
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533:452–454
- Barbui C (2016) Sharing all types of clinical data and harmonizing journal standards. *BMC Med* 14:63
- Berghmans S et al (2017) Open Data: the researcher perspective – survey and case studies. <https://data.mendeley.com/datasets/bwrnfb4bv/1>
- Björk B-C (2015) Have the “mega-journals” reached the limits to growth? *PeerJ* 3:e981
- Blohowiak BB (2013) Badges to acknowledge open practices. https://osf.io/tvyxz/files/?_ga=2.252581578.297610246.1542300800-587952028.1539080384
- Bucci EM (2018) Automatic detection of image manipulations in the biomedical literature. *Cell Death Dis* 9:400

- Burton A et al (2017) The Scholix framework for interoperability in data-literature information exchange. *D-Lib Mag* 23:1
- Butler D (2000) BioMed central boosted by editorial board. *Nature* 405:384
- Callaghan S et al (2014) Guidelines on recommending data repositories as partners in publishing research data. *Int J Digit Curation* 9:152–163
- Colavizza et al (2019) The citation advantage of linking publications to research data. <https://arxiv.org/abs/1907.02565>
- Chan A-W et al (2017) Association of trial registration with reporting of primary outcomes in protocols and publications. *JAMA* 318:1709–1711
- Cobo E et al (2007) Statistical reviewers improve reporting in biomedical articles: a randomized trial. *PLoS One* 2:e332
- Cousijn H et al (2017) A data citation roadmap for scientific publishers. *BioRxiv*. <https://doi.org/10.1101/100784>
- Cousijn H et al (2018) A data citation roadmap for scientific publishers. *Sci Data* 5:180259
- Dorch BF, Drachen TM, Ellegaard O (2015) The data sharing advantage in astrophysics. *Proc Int Astron Union* 11:172–175
- Emerson GB et al (2010) Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Arch Intern Med* 170:1934–1939
- Eyding D et al (2010) Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 341:c4737
- Fang FC, Steen RG, Casadevall A (2012) Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A* 109:17028–17033
- Ferguson AR, Nielson JL, Cragin MH, Bandrowski AE, Martone ME (2014) Big data from small data: data-sharing in the “long tail” of neuroscience. *Nat Neurosci* 17:1442–1447
- Freedman LP, Cockburn IM, Simcoe TS (2015) The economics of reproducibility in preclinical research. *PLoS Biol* 13:e1002165
- Garza K, Fenner M (2018) Glad you asked: a snapshot of the current state of data citation. <https://blog.datacite.org/citation-analysis-scholix-rda/>
- Giofrè D, Cumming G, Fresc L, Boedker I, Tressoldi P (2017) The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. *PLoS One* 12:e0175583
- Goldacre B et al (2018) Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. *BMJ* 362:k3218
- Graf C (2018) How and why we're making research data more open. Wiley, Hoboken. <https://www.wiley.com/network/researchers/licensing-and-open-access/how-and-why-we-re-making-research-data-more-open>
- Grant R, Smith G, Hrynaszkiewicz I (2019) Assessing metadata and curation quality: a case study from the development of a third-party curation service at Springer Nature. *BioRxiv*. <https://doi.org/10.1101/530691>
- Grieneisen ML, Zhang M (2012) A comprehensive survey of retracted articles from the scholarly literature. *PLoS One* 7:e44118
- Hardwicke TE et al (2018) Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*. *R Soc Open Sci* 5:180448
- Hrynaszkiewicz I, Cockerill MJ (2012) Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. *BMC Res Notes* 5:494
- Hrynaszkiewicz I, Shintani Y (2014) Scientific Data: an open access and open data publication to facilitate reproducible research. *J Inf Process Manag* 57:629–640
- Hrynaszkiewicz I, Busch S, Cockerill MJ (2013) Licensing the future: report on BioMed Central's public consultation on open data in peer-reviewed journals. *BMC Res Notes* 6:318
- Hrynaszkiewicz I, Li P, Edmunds SC (2014) In: Stodden V, Leisch F, Peng RD (eds) *Implementing reproducible research*. CRC Press, Boca Raton

- Hrynaszkiewicz I, Khodiyar V, Hufton AL, Sansone S-A (2016) Publishing descriptions of non-public clinical datasets: proposed guidance for researchers, repositories, editors and funding organisations. *Res Integr Peer Rev* 1:6
- Hrynaszkiewicz I et al (2017a) Standardising and harmonising research data policy in scholarly publishing. *Int J Digit Curation* 12:65
- Hrynaszkiewicz I, Simons N, Goudie S, Hussain A (2017b) Research Data Alliance Interest Group: data policy standardisation and implementation. <https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation>
- Hrynaszkiewicz et al (2019) Developing a research data policy framework for all journals and publishers. <https://doi.org/10.6084/m9.figshare.8223365.v1>
- I4OC: initiative for open citations. <https://i4oc.org/>
- Inchcombe S (2017) The changing role of research publishing: a case study from Springer Nature. *Insights* 30:13–19
- Ioannidis JPA et al (2009) Repeatability of published microarray gene expression analyses. *Nat Genet* 41:149–155
- Johnson R, Focci M, Chiarelli A, Pinfield S, Jubb M (2017) Towards a competitive and sustainable open access publishing market in Europe: a study of the Open Access Market and Policy Environment. *OpenAIRE*, Brussels, p 77
- Kidwell MC et al (2016) Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Biol* 14:e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
- Kirk R, Norton L (2019) Supporting data sharing. *NPJ Breast Cancer* 5:8
- Kovanis M, Porcher R, Ravaud P, Trinquart L (2016) The global burden of journal peer review in the biomedical literature: strong imbalance in the collective enterprise. *PLoS One* 11:e0166387
- Leonelli S (2016) Open data: curation is under-resourced. *Nature* 538:41
- Lewis LM et al (2018) Replication study: transcriptional amplification in tumor cells with elevated c-Myc. *Elife* 7
- Li T et al (2016) Review and publication of protocol submissions to trials – what have we learned in 10 years? *Trials* 18:34
- Lin J (2018) Preprints growth rate ten times higher than journal articles. <https://www.crossref.org/blog/preprints-growth-rate-ten-times-higher-than-journal-articles/>
- Luther J (2017) The stars are aligning for preprints – the scholarly kitchen. <https://scholarlykitchen.sspnet.org/2017/04/18/stars-aligning-preprints/>
- Macleod MR, The NPQIP Collaborative Group (2017) Findings of a retrospective, controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the completeness of reporting study design and execution. *BioRxiv*. <https://doi.org/10.1101/187245>
- Markowitz F (2015) Five selfish reasons to work reproducibly. *Genome Biol* 16:274
- McGauran N et al (2010) Reporting bias in medical research – a narrative review. *Trials* 11:37
- McNutt M (2014) Journals unite for reproducibility. *Science* 346:679
- Michener WK (2015) Ten simple rules for creating a good data management plan. *PLoS Comput Biol* 11:e1004525
- Munafò MR et al (2017) A manifesto for reproducible science. *Nat Hum Behav* 1:0021
- Nature (2015) Ctrl alt share. *Sci Data* 2:150004
- Nature (2017) Extending transparency to code. *Nat Neurosci* 20:761
- Nature (2018) Checklists work to improve science. *Nature* 556:273–274
- Naughton L, Kernohan D (2016) Making sense of journal research data policies. *Insight* 29:84–89
- Nosek BA, Lakens D (2014) Registered reports. *Soc Psychol* 45:137–141
- Nosek B et al (2014) Transparency and openness promotion (TOP) guidelines. https://osf.io/xd6gr/?_ga=2.251468229.297610246.1542300800-587952028.1539080384
- Nuijten MB et al (2017) Journal data sharing policies and statistical reporting inconsistencies in psychology. *Collabra Psychol* 3:31
- Peng RD (2009) Reproducible research and biostatistics. *Biostatistics* 10:405–408

- Percie du Sert N et al (2018) Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Sci* 2:e000002
- Pienta AM, Alter GC (2010) The enduring value of social science research: the use and reuse of primary research data. *Russell*. <http://141.213.232.243/handle/2027.42/78307>
- Piwowar H (2013) Altmetrics: value all research products. *Nature* 493:159
- Piwowar HA, Vision TJ (2013) Data reuse and the open data citation advantage. *PeerJ* 1:e175
- Piwowar HA, Day RS, Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. *PLoS One* 2:e308
- Polanin JR, Terzian M (2018) A data-sharing agreement helps to increase researchers' willingness to share primary data: results from a randomized controlled trial. *J Clin Epidemiol* 106:60–69
- Prinz F, Schlange T, Asadullah K (2011) Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 10:712
- Rathi V et al (2012) Clinical trial data sharing among trialists: a cross-sectional survey. *BMJ* 345:e7570
- Robertson M (2017) Who needs registered reports? *BMC Biol* 15:49
- Rowhani-Farid A, Barnett AG (2016) Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ Open* 6:e011784
- Rowhani-Farid A, Barnett AG (2018) Badges for sharing data and code at biostatistics: an observational study. [version 2; referees: 2 approved]. *F1000Res* 7:90
- Rowhani-Farid A, Allen M, Barnett AG (2017) What incentives increase data sharing in health and medical research? A systematic review. *Res Integr Peer Rev* 2:4
- Savage CJ, Vickers AJ (2009) Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* 4:e7078
- Schmidt B, Gemeinholzer B, Treloar A (2016) Open data in global environmental research: the Belmont forum's open data survey. *PLoS One* 11:e0146695
- Science D et al (2017) The state of open data report 2017. Digital Science, London
- Scientific Data (2019) Scientific Data recommended repositories. https://figshare.com/articles/Scientific_Data_recommended_repositories_June_2015/1434640
- SciGraph | For Researchers | Springer Nature. <https://www.springernature.com/gp/researchers/scigraph>
- Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF (2016) Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. *Trials* 17:301
- Shotton D (2013) Publishing: open citations. *Nature* 502:295–297
- Simera I et al (2010) Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 8:24
- Smith R (2010) Classical peer review: an empty gun. *Breast Cancer Res* 12(Suppl 4):S13
- Smith G, Grant R, Hrynaszkiwicz I (2018) Quality and completeness scores for curated and non-curated datasets. https://figshare.com/articles/Quality_and_completeness_scores_for_curated_and_non-curated_datasets/6200357
- Springer Nature (2018) Springer Nature launches Open data badges pilot – Research in progress blog. <http://blogs.biomedcentral.com/bmcblog/2018/10/08/springer-nature-launches-open-data-badges-pilot/>
- STM, ALPSP (2006) Databases, data sets, and data accessibility – views and practices of scholarly publishers. https://www.stm-assoc.org/2006_06_01_STM_ALPSP_Data_Statement.pdf
- Stuart D et al (2018) Whitepaper: practical challenges for researchers in data sharing. https://figshare.com/articles/Whitepaper_Practical_challenges_for_researchers_in_data_sharing/5975011
- Suber P (2012) Open access. MIT Press, Cambridge. http://mitpress.mit.edu/sites/default/files/titles/content/9780262517638_sch_0001.pdf
- Taichman DB et al (2017) Data sharing statements for clinical trials. *BMJ* 357:j2372
- Teixeira da Silva JA (2015) Negative results: negative perceptions limit their potential for increasing reproducibility. *J Negat Results Biomed* 14:12

- Text and Data Mining | Springer Nature | For Researchers | Springer Nature. <https://www.springernature.com/gp/researchers/text-and-data-mining>
- Text and Data Mining Policy – Elsevier. <https://www.elsevier.com/about/policies/text-and-data-mining>
- The NPQIP Collaborative Group (2019) Did a change in Nature journals’ editorial policy for life sciences research improve reporting? *BMJ Open Sci* 3:e000035
- UK Research and Innovation (2011) Common principles on data policy. <https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>
- Vanpaemel W, Vermorgen M, Deriemaecker L, Storms G (2015) Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra* 1
- Vasilevsky NA, Minnier J, Haendel MA, Champieux RE (2017) Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ* 5:e3208
- Vines TH et al (2013) Mandated data archiving greatly improves access to research data. *FASEB J* 27:1304–1308
- Whitlock MC, McPeck MA, Rausher MD, Rieseberg L, Moore AJ (2010) Data archiving. *Am Nat* 175:145–146
- Wicherts JM, Borsboom D, Kats J, Molenaar D (2006) The poor availability of psychological research data for reanalysis. *Am Psychol* 61:726–728
- Wiley Open Science Researcher Survey (2016) https://figshare.com/articles/Wiley_Open_Science_Researcher_Survey_2016/4748332/2

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

