# Chapter 6
# Open Access and Research Reproducibility in Biomedical Sciences

**Shimelis Getu Assefa**
*University of Denver, USA*

## ABSTRACT

*Reproducibility-enhancing practices of open access journals in biomedical sciences are investigated. Based on transparency and openness promotion guidelines and relevant reporting requirements by institutions that are in the forefront of advancing reproducibility research, eight standards were used to evaluate 27 biomedical journals to 1) determine the extent to which these journals address reproducibility, 2) identify specific policy themes required, and 3) understand overall infrastructure promoted by the journals to deposit, archive, share, and discover research assets. The results show that almost all the 27 journals required authors to address six of the eight standards when preparing and submitting their research. Two standards that were not frequently addressed are preregistration of the study and preregistration of analysis plans. 'Data availability' policy is the most recurring theme across all journals. The infrastructure promoted to manage the overall scholarly communication workflow range from data, code, software repositories, protocol registration, to funding registry.*

## INTRODUCTION

Open Access (OA) is becoming an established method of scholarly communication. After 20 years of application and experimentation - social, technical, political, and economic factors have converged to make OA a viable form of knowledge production and dissemination. While there were attempts to archive preprints in the early 1980s in ftp servers such as the arXiv project, it's by the turn of the 21st century that open access movement gained momentum after the publishing of manifestos and initiatives such as the Budapest Open Access Initiative (BOAI) (Harnad, et al., 2004), Bethesda Statement on Open Ac-

DOI: 10.4018/978-1-7998-9805-4.ch006

cess Publishing (Brown, et al., 2003), and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (Redalyc, 2003).

Open Access (OA) is defined as "access to literature that is digital, Online, free of charge, and free of most copyright and licensing restrictions," (Suber, 2010). This definition places emphasis on full open access and hides so much details as to what, when, and how the scholarly literature is to be made available in the open. Open access publishing is achieved, with some slight variations, in two ways: gold open access and green open access (Gargouri, 2012; Harnad, et al., 2004). In Gold OA model (mainly in OA Journals), published works are made freely accessible to readers after the cost of the article processing is borne by the author, or by a sponsoring agency, and in Green OA, while there are different variations, publishers allow authors to self-archive their work (Harnad, et al., 2008).

On the other hand, research reproducibility, or the idea of making one's research reproducible garners different definitions and accounts across the scientific enterprise. Often different terms such as - reproducibility, replicability, repeatability, reliability, robustness, and generalizability – are used and create confusion (Goodman et al., 2016). While remaining within the umbrella of 'reproducible research,' extensive coverage is given in the extant literature to the kinds of reproducibility investigated such as – methods reproducibility, results reproducibility, and inferential reproducibility (Goodman et al., 2016) and computational reproducibility (e.g., Grüning et al, 2018). In this chapter, research reproducibility is broadly looked at that adheres to reporting guidelines on data, methods, statistical analytical steps, computational analysis and software codes, and overall research design.

Most importantly the focus of this chapter is to investigate the degree to which open access scholarly journals promote transparency and reproducibility-enhancing practices. The fact that open access journals are shared freely and in the open on the Internet, it is easy to verify the steps taken by the original study so other independent researchers can achieve similar findings, of course within the confines of the new research parameters. The question, however, is which of the bio-medical sciences open access journals require that authors submit manuscripts alongside data, software codes, and sufficient documentation of the overall research design to promote reproducibility-enhancing practices.

## BACKGROUND

Overall, there is an increasing recognition of the fact that sharing published works in the open Internet is vital for scientific progress. In addition to the established scholarly journals that follow either the Green or Gold model, the OA distribution ecosystem is fast expanding that covers - for example personal websites (e.g., https://terrytao.wordpress.com/) to OA Institutional Repositories (e.g., https://deepblue.lib.umich.edu/documents) to aggregators, discovery and social reference management systems such as Mendeley (Thelwall, 2018), to OA Directory of Journals (DOAJ) (Morrison, 2017), to research work flow management systems (e.g., https://galaxyproject.org/ & https://wholetale.org/) to repositories of preprints and postprints (e.g., arXiv, bioRxiv, medRxiv, PsyArXiv), to fully OA academic & scholarly publishers (e.g, https://plos.org/), and to OA software - Essential Open Source Software for Science (EOSS) (e.g., https://chanzuckerberg.com/eoss/).

According to ULRICH'S database (Ulrichsweb), the most comprehensive database of periodicals with international coverage (English and non-English), there are about 89,000 active academic and scholarly journals that are referred and/or peer-reviewed - as shown by the search result using the following syntax:

(89,039 results for: Status:("Active") Serial Type:("Journal") Content Type:("Academic / Scholarly") Key Feature:(+"Refereed / Peer-reviewed").

Further refining this search syntax to retrieve OA scholarly and academic journals, about 22,000 results are returned as shown by the following search syntax:

(22,443 results for: Status:("Active") Serial Type:("Journal") Content Type:("Academic / Scholarly") Key Feature:(+"Refereed / Peer-reviewed" +"Open Access").

Without going into a detailed analysis of the tens of thousands of OA scholarly journals returned by the search results above, it is safe to state that about a quarter (24.7%) of the global peer-reviewed journals are now available under some version of open access model - a sign that OA academic and scholarly journals have entered a critical mass.

On the other hand, in the larger context of open science movement, reproducibility has gained momentum to promote openness and transparency in research findings by sharing research and associated assets in publicly accessible outlets. An even more important impetus for reproducibility gaining momentum is because significant proportion of research dollar has gone wasted - as for example in the rate of failure of clinical trials (e.g., Bøtker, et al., 2018), safety issues with approved drugs (e.g., Kannt & Wieland, 2016), or in general reproducibility crisis in published research (Bolli, 2017; França & Monserrat, 2018).

Reproducibility is about rigor, transparency, openness, validation, corroboration, and integrity of published studies. According to the U.S. National Science Foundation (NSF) subcommittee on replicability in science, "reproducibility refers to the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator (National Academies of Sciences, Engineering, and Medicine, 2019). In other words, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis to yield the same results, hence reproducibility is a minimum necessary condition for a finding to be believable and informative (National Academies of Sciences, Engineering, and Medicine, 2019).

While it is easier to recognize the power of reproducibility in the ethical conduct of research, it is often difficult to achieve it in the overall scholarly communication domain. An empirical review of existing literature reveals that there is a consensus about the prevalence of 'reproducibility crisis,' in scientific research (e.g., "Announcement", 2017). According to a survey of 1,576 scientists by Nature 52% agree that there is a significant crisis, and 38% agree to the presence of a slight crisis in reproducibility research, bringing the total to close to 90% (Baker, 2016). The same survey showed that based on 185 research participants from Nature Research, they indicated the major issues for irreproducibility as: selective reporting, pressure to publish, low statistical power or poor analysis, not replicated enough in original lab, insufficient oversight/mentoring, methods, code unavailable, poor experimental design, raw data not available from original lab, fraud, and insufficient peer review (Baker, 2016).

## Existing Standards and Guidelines for Enhancing Reproducibility

In biological, medical, and scientific communities, there are several initiatives to develop standards and guidelines to enhance reproducibility in scholarly communications. This chapter reviewed the most important ones to offer context regarding the study at hand.

The National Academies of Sciences, Engineering, Medicine, has published a Consensus Study Report titled "Reproducibility and Replicability in Science" (National Academies of Sciences, Engineering, and Medicine, 2019), in which it included numerous examples of reports designed to improve research practices with a goal of improving reproducibility and replicability. One of the reports by the Federation of American Societies for Experimental Biology (FASEB, 2016), recommend principles and best practices of openness, reproducibility, and rigor in data and science.

Another effort in the biological and medical sciences comes from BioMed Central – a Springer Nature affiliate that publishes about 300 peer-reviewed journals in science, technology, and medicine – developed what it called 'BioMed Central Minimum Standards of Reporting Checklist,' (Kenall et al, 2015). The checklist addresses three areas, i.e., experimental design and statistics, resources, and availability of data and materials.

A similar and most robust guideline is published by the Center for Open Science that is known as Transparency and Openness Promotion (TOP) Guidelines (Center for Open Science, 2021). TOP Guidelines covers eight transparency and openness standards: Citation; Data transparency; Analytic methods (code) transparency; Research materials transparency; Design and analysis transparency; Preregistration of studies; Preregistration of analysis plans; and Replication.

## THESIS STATEMENT

As noted, open access (OA) scholarly and academic journals are gaining wider acceptance as one model of research outlets in the larger scheme of scholarly communication. Equally, there is a growing call in the scientific enterprise for transparent and ethical research practices that promote. In view of that, this chapter seeks to investigate the degree to which scientific OA journals in biological and medical sciences promote and advance reproducibility-enhancing practices. What are some of the requirements placed by OA Journals to support reproducibility practices?
Most importantly, this chapter seeks to investigate the following research questions:

1. To what extent open access scholarly journals in Biomedical Sciences promote reproducibility-enhancing research standards?
2. What are the standards in reporting requirement specified by OA Scholarly journals in Biomedical Sciences?
3. What types of infrastructure are built or promoted by OA Biomedical scholarly journals to support reproducibility-enhancing documentation?

### Statement of Purpose

The purpose of this study is to analyze the reporting requirements and guidelines of OA biological and medical science referred and peer-reviewed scholarly journals. By analyzing the text in the author guidelines for manuscript submission, this study aims to make an empirical contribution on how OA scholarly journals address reproducibility-enhancing practices.

## METHODOLOGY

Ulrich's International Periodicals Directory (Ulrichsweb.com) is one of the single largest global authoritative databases that is used to identify serials and periodicals in a wide range of fields, including peer-reviewed and referred journals. Ulrichsweb is an easy to search source of detailed information on more than 300,000 periodicals (also called serials) of all types: academic and scholarly journals, e-journals, peer-reviewed titles, popular magazines, newspapers, newsletters, and more (Meeks, 2018).

Using Ulrich's database, a search was made to identify and retrieve open access scholarly journals in Biological Sciences. For subject areas, Ulrich's database treat 'Biological Sciences and Agriculture' together and that is the subject category selected. Accordingly, the following search syntax produced 186 active open access scholarly journals in the Biological Sciences and Agriculture. English language is selected as the language of text for the journals retrieved, and the Journal Citation report is also checked as a filter to retrieve highly rated journals. The following is the syntax of the search and number of results returned:

*186 results for: Status:("Active") Serial Type:("Journal") Content Type:("Academic/Scholarly") Subject Areas:("Biological Sciences and Agriculture") Key Feature:(+"Refereed/Peer-reviewed" +"Available Online" +"Open Access" +"Journal Citation Reports") Language of Text:("English")*

The search results were exported into Excel spreadsheet to further filter the top Biology (all aspects from Botany to Zoology) journals that are flagged both 'referred' and 'peer-reviewed' and that resulted in 27 journals. The Appendix summarizes the 27 journals selected for the analysis.

The website of each of the 27 open access journals were visited to extract the text provided in the 'Information for Authors' or 'Guide for Authors' section focusing on the 'materials and methods', 'preparing your manuscript', and any section that addresses transparency and reproducibility-enhancing practices. Assumption is made that authors will follow guidelines and instructions provided while preparing their manuscripts.

In order to find common baseline criteria, in addition to respective journals guidelines and reporting requirements, attempt is also made to review guidelines by institutions that are in the forefront of advancing reproducibility research, such as the NIH Principles and Guidelines for Reporting Preclinical Research https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research. As a result, one more requirement, COI or competing interests, is added to the 'Guidelines for Transparency and Openness Promotion (TOP) in Journal Policies and Practices', or 'The Top Guidelines' in short was used (Center for Open Science, 2021) to assess transparency and reproducibility in the 27 selected OA biological science journals.

The Top Guidelines has eight transparency standards, but only the first seven are used for this study as the eighth standard is applicable for replication research that's beyond the scope of this study. As noted above, from other reporting guidelines and requirements, conflict of interest or competing interests is added as one criterion to evaluate reproducibility-enhancing practices. Hence, a synthesized summary of the eight standards is provided below:

1. Citation - citation of articles is routine and well-formulated. Similar standards can be applied to citation of data, code, and materials to recognize and credit these as original intellectual contributions.

2. Data transparency - for published articles, whether or not data is available and, if available, how to access it, or data must be posted to a trusted repository, or exceptions to sharing data for legal or ethical reasons must be stated to the editor at the time of submission.
3. Analytic methods (code) transparency - for published articles, whether or not program code is available and, if available, how to access it, or provide program code, scripts, codebooks, and other documentation sufficient to precisely reproduce all published results or provide software and other documentation that will precisely reproduce all published results.
4. Research materials transparency - for published articles, whether or not research materials are available and, if available, how to access it, or materials used to conduct the research are clearly and precisely documented and are maximally available to any researcher for purposes of reproducing the results or replicating the procedure, or all materials supporting the claims made by the author must be made available to the journal prior to publication, exceptions to sharing materials for legal or ethical reasons must be stated at the time of submission.
5. Design and analysis transparency - standards for reporting research design and analysis should maximize transparency about the research process and minimize potential for vague or incomplete reporting of the methodology.
6. Preregistration of studies - is a means of making research more discoverable even if it does not get published. By encouraging or requiring preregistration, journals increase the likelihood of discoverability of research that is not ultimately published.
7. Preregistration of analysis plans - certifies the distinction between confirmatory and exploratory research. Preregistration of analysis plans supersedes preregistration of studies.
8. Conflict of Interest, or, competing interests – disclosing any actual or perceived conflicts of interest or financial interests on the part of author/s that might raise the question of bias in the work reported.

## RESULTS AND DISCUSSION

### Results

To answer the research questions, the text from the 'instructions for authors' or 'Guides for Authors' of the 27 OA peer-reviewed journals were individually examined to determine the extent to which the journals promote reproducibility-enhancing practices; to identify specific reporting requirements; and to understand allied infrastructure developed by the journals or third-party tools & technologies used to document, preserve, and share artifacts accompanying published works.

**RQ#1:** To what extent open access scholarly journals in Biological Sciences promote reproducibility-enhancing research standards?

The grid below (Table 1) presents data collated from each of the 27 OA journals evaluated based on the 8 standards to determine whether the journals require the reporting standards. The eight standards correspond to the list provided above in the 'Methodology' chapter. In the grid '1' means Yes, the Journal requires authors to meet those standards; and '0' means No the journal doesn't require the corresponding standards.

**RQ#2:** What are the standards in reporting requirement specified by OA Scholarly journals in Biological Sciences?

*Table 1. Top biomedical OA journals and their requirements to promote reproducibility-enhancing practices*

| Journal Name | Standard #: | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| AOB Plants | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| Biology Open | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BioMed Central Bioinformatics | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Biology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Biotechnology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Genomic Data | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Genomics | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Mol. and Cell Biology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Plant Biology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| BMC Zoology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Cell Discovery | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EMBO Molecular Medicine | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| European Jnl of Cell Biology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Frontiers in Plant Science | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Frontiers in Zoology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Genome Biology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Jnl of Biological Chemistry | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Journal of Marine Sciences | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Journal of Plant Interactions | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Marine Biodiversity Records | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Microbiology Open | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| New Disease Reports | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| Nucleic Acids Research | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Open Biology | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Plant Biotechnology Journal | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| PLoS Biology | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| PLoS Computational Biology | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Across the 27 journals selected in this study, the text from the section on 'Instructions for Authors,' or 'Guides for Authors' is examined and coded to determine the most frequently occurring themes used to address each of the eight standards used in this study. The list below presents the most common standard statements required by the journals for submission and reporting purposes:

**On Citation**

- We have endorsed the FORCE11 Data Citation Principles. Data must be cited in the same way as article, book, and web citations and authors are required to include data citations as part of their reference list.
- Data citations should include a persistent identifier (such as a DOI) and should ideally be included in the reference list.
- Citations of datasets, when they appear in the reference list, should include the minimum information recommended by DataCite and follow journal style.

**On Data Availability**

- Must indicate where the data described in the manuscript are located. For datasets that were deposited into a publicly accessible repository, the location and identifying information (i.e., accession numbers or DOIs) must be provided.
- Authors must provide access to the data underlying the results presented in their article.
- All datasets on which the conclusions rely should be either deposited in publicly available repositories (where available and appropriate) or presented in the main manuscript or additional supporting files whenever possible.
- All data and related metadata underlying the findings reported in a submitted manuscript should be deposited in an appropriate public repository, unless already provided as part of the submitted article.
- The data should be FAIR—findable, accessible, interoperable, and reusable—so that other researchers can locate and use the data.

**On Materials Availability**

- If the study characterizes the activity of new compounds, compound structures and the protocol for obtaining the compound must be provided.
- Materials and Methods descriptors - e.g., reagents including enzymes, antibodies, kits, specialized commercial instruments, non-standard chemicals, peptides, recombinant proteins, et al. with reference to company name, location, and catalog # must be provided.
- Explicitly requires that authors will provide, for non-profit research, all the biological and chemical materials not commercially available, including all plant cultivars, cell lines, DNA, antibodies, and other similar materials, that are used for the experiments reported.

**On Code, Program, Software, Statistical Analysis**

- Software code should be archived in a repository that can assign it a DOI and the DOI should be provided.
- Make the program accessible as a web server with no login requirements, deposit the source code in a public repository such as GitHub, or upload an executable version of the program and instructions as supplementary data.
- Software must be freely available to users at the time of submission, either as executable versions for multiple, common platforms (Linux, Windows and MacOS) or as source code or as a web server.
- Statistical Analyses: Details to mathematical evaluations for quantifying accuracy, sensitivity, and reproducibility should be included, indicating equation choice and numbers of replicates.
- Source code for any in-house scripts that are used for analyses must also be either deposited in a public repository or included in the supplementary materials.

- Provide all code used to generate statistics & generate figures, along with any (processed) data required as inputs, along with details of what software it requires (program and version)

**On Design, Methods, Protocols**

- Describe the overall experimental design briefly but with sufficient information to permit a qualified reader to repeat the experiments.
- Include the DOI link in the Methods section of your manuscript using the following format provided by protocols.io: https://dx.doi.org/10.17504/protocols.io.[PROTOCOL DOI]
- Experimental procedure described in a manuscript would benefit from a step-by-step protocol, submit detailed protocols for peer review and publication in Bio-protocol or to deposit them at protocols.io.
- In exceptional cases where the materials and methods are particularly lengthy, more detailed experimental protocols or descriptions of computational analyses may be included as supplementary materials and methods (though all methods must be mentioned in the main text).

**On Pre-Registration of Studies**

- You should always ensure that you register protocols for clinical trials you're involved with in a publicly accessible registry.
- Authors are encouraged to indicate whether the conducted research was preregistered in an independent, institutional registry.
- Preregistration of studies involves registering the study design, variables, and treatment conditions prior to conducting the research.

**On Pre-Registration of Analysis**

- Authors are encouraged to indicate whether or not the conducted research was preregistered with an analysis plan in an independent, institutional registry.
- Preregistration of analysis plans include specification of sequence of analyses or the statistical model that will be reported.
- Nature Portfolio journals support pre-registration of analysis plans in public repositories; details of pre-registration should be provided with submission.

**On COI, or Competing Interests**

- In the interests of transparency and to help readers form their own judgments of potential bias, corresponding authors are required to declare any competing financial and/or non-financial interests in relation to the work described.
- Must include a statement disclosing whether there are any actual or perceived conflicts of interest on the part of any author.
- In the submission system, declare on behalf of all authors whether there are any financial, personal, or professional interests that could be construed to have influenced the work.

**RQ#3:** What types of infrastructure are built or promoted by open access scholarly journals to support reproducibility-enhancing documentation?

While most of the journals don't have the required Infrastructure built within their publishing workflow, they require or recommend authors to link to third-party tools and technologies where research assets are deposited, shared, and archived, and documented. These research assets are critical to support the management, publishing, archiving, sharing, discovering, version control, and overall documentation of the scholarly communication process, and hence enhance reproducibility. Representative infrastructure that the journals recommend include the following:

- For protocol registration - Bio-protocol, Protocol.io, International Clinical Trials Registry Platform (ICTRP)
- For OA license - CC-BY Creative Commons attribution license
- For research registration -Crossmark for clinical trials, OSF.io, clinicaltrials.gov, Crossref
- For publicly available data repositories - Zenodo, figshare, GenBank, Mendeley Data
- For depositing nucleotide sequence data - DDBJ, EMBL, or GenBank
- For depositing confidential human data – dbGaP, dbSNP, European Genome-phenome Archive EGA
- For persistent data, protocol, and research identifier - DOI
- For data citation - FORCE11, DataCite
- For code, software archiving / sharing - GitHub, Zenodo, CodeOcean, Software Heritage archive
- For COI, funding registry - Open Funder Registry Crossref funder registry
- For authors identification - ORCID

## DISCUSSION

From the data presented above it is evident that almost all the 27 journals investigated in this study support reproducibility enhancing practices. The only 2 standards that are sparsely addressed are: 1) Pre-registration of studies; and 2) Pre-registration of analysis plans. The emphasis on 'data availability' requirement by these journals confirms the prevailing consensus that 'no data, no science' (Miyakawa, 2020) as key requirements for enhancing reproducibility.

**RQ#1:** To what extent open access scholarly journals in Biological Sciences promote reproducibility-enhancing research standards?

Analysis of the reporting requirements text from the 'Instructions for Authors' or 'Guides for Authors' sections of the 27 journals reveal that while six of the eight reproducibility-enhancing standards are clearly specified, two standards, i.e, 1) preregistration of study, and 2) preregistration of analysis plans; were addressed only in 5 of the 27 journals. The five journals that address all the eight standards are: 1) Cell Discovery, 2) Journal of Marine Science, 3) Journal of Plant Interactions, 4) PLoS Biology, and 5) PLoS Computational Biology. In addition, strong emphasis is placed by all journals on 'data availability' as a must for submission. The kinds of infrastructure suggested and promoted by these journals as part of the scholarly communication workflow range from data, code, software repositories, protocol registration, to funding registry. There is no single journal that has built a complete infrastructure for all research assets in its publishing workflow system. The journals usually enhance submitted reports by adding metadata and linking to various endpoints to their-party tools and technologies.

While individual variations exist across the 27 journals in terms of how exhaustively they address the submission requirements, it is evident that almost all open access biological sciences journals investigated in this study require reproducibility-enhancing practices. This is also a testament to how far open access journals have come in the overall scholarly communication ecosystem.

**RQ#2:** What are the standards in reporting requirement specified by OA Scholarly journals in biological sciences?

In the 'Information for Authors', or 'Guides for Authors' sections of Author Guidelines in each of the 27 selected journals, specific attention is given to the text under 'preparing your manuscript' sections with particular focus on 'data availability', 'materials and methods', 'experimental procedures', 'research design', and 'funding disclosure or competing interests' sections of the journals. Some of the journals promote stringent requirements and offer detailed guidelines, while others require the minimum standards necessary to interpret the findings of the study. At a minimum almost all journals require data availability and sufficient description of research design.

On the other spectrum, for example the 'Journal of Plant interactions,' provides robust guidelines starting from - Declarations and ethics statements- where prior to starting the study, ethical approval must be obtained, and that all studies involving humans are performed in accordance with the principles stated in the Declaration of Helsinki, and (when applicable) informed consent from participants is voluntarily obtained in accordance with the principles outlined in the Nuremberg Code, the Belmont Report and the American Anthropological Association.

Another example that demonstrates stringent and detailed guideline is found in the journal 'Nucleic Acid Research (NAR).' In NAR, for example, when depositing data involving human participants, authors must ensure that all datasets have been de-identified and anonymized in accordance with the Safe Harbor method before submission. If needed, a trusted third party can be used to convert personal data into anonymized data, or NAR recommends appropriate databases for depositing confidential human data. In general, other than lack of policy and standard guideline in the two areas, i.e., 1) preregistration of studies; and 2) preregistration of analysis plans; all the 27 journals have clear statements that require authors to address reproducibility-enhancing practices on the remaining six areas: data citation; data availability; material availability; sharing code, program, software, and statistical analysis; details of research design, methods, and protocol; and disclosure of conflict of interest.

In addition to a must have 'data availability' policy, the research design, materials, and methods section requirement is given great significance in reproducibility-enhancing research. When it comes to the overall research design, the standard requirements by many of the journals address standards such as - detailed reporting instructions of actual experimental design including: the exact sample size (n) for each experimental group/condition; a precise description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, laboratory replicates, etc.); if comparisons are made between groups, a clear description of the method used for statistical inference; for the analysis of categorical outcomes, exact methods or asymptotic methods with appropriate adjustments for small sample size; standard chi-squared or difference in proportions tests for large sample sizes; t-tests for comparisons of continuous data that are normally distributed data to compare two groups; Analyses of Variance (ANOVA) for comparisons between three or more groups; and If p-values are presented, one-sided or two-sided should be specified, and if one-sided, justification should be provided.

**RQ#3:** What types of infrastructure are built or promoted by open access scholarly journals to support reproducibility-enhancing documentation?

To support documentation and sharing of research assets such as data, codes, protocols, etc., the biological sciences journals sampled in this study recommend tools and technologies (mostly third-party, outside of the respective publishing platforms of the journals). The journals then provide links to the various endpoints of the research assets for easier identification and discovery. This can be often

challenging for authors to jump from one system to another to deposit and share the different research assets from data to codes/scripts, to material, and to protocol, and funding registry. It would be easier and more streamlined if journal publishers provide a one-stop platform during the manuscript submission process for authors to weave and compile the manuscript and all accompanying assets that are required to promote reproducibility. The goal should be towards automated submission process that also ensures long-term preservation of the different artifacts of the research work. In this regard assigning persistent or permanent identifier such as DOI for the different research assets, as well as the proper documentation and adequate description of data, materials, statistical methods, analysis, scripts, software, are critical to further enhance reproducibility research.

## FUTURE RESEARCH DIRECTIONS

There is no question reproducibility-enhancing research practices are critical for the advancement of science and scientific knowledge, in general. This chapter showed open access scholarly journals in biomedical sciences, to a large extent, have built reporting requirements that adhere to reproducibility-enhancing practices. However, the overall workflow, from pre-registration to submission, is made up of independent and third-party tools and technological infrastructure outside of the specific journal manuscript submission system. To hop from one system to another to document analysis steps, register protocols or conflict of interest, and/or deposit associated research assets such as data is often daunting for researchers that can easily result in less-optimal documentation and reporting standards that in the end will have adverse effects on reproducibility.

To minimize the additional time and effort required of researchers to comply to reproducibility-enhancing research practices, it is prudent if journal publishing systems investigate a one-stop platform that supports the overall research workflow from start to completion. This could mean easy porting of manuscripts author computer to publication submission system, or built-in software to author manuscripts, codes, analysis, and easy linking to additional research assets that in the end can be compiled to generate a well documented report. In addition to time and effort, it is important to consider the long-term preservation and version control dimensions of published works.

## CONCLUSION

The implications of this study are in advancing the synergy between open access publications and reproducibility-enhancing research practices. Both open access and reproducibility share overlapping goals, i.e., wider dissemination, openness, verification, transparency, and establishing the integrity and ethical conduct of research, especially those involving human subjects. Outside of open access scholarly communication, it is equally important if journals and scholarly publications behind paywalls consider to promote reproducibility by opening access to critical research artifacts such as data, software, code, and analyses, for rigorous verifications.

Top open access biomedical sciences journals are investigated to determine the extent to which their reporting requirements advance reproducibility. These reporting requirements are evaluated against the most comprehensive reproducibility-enhancing standards developed by the Center for Open Science called ‘Guidelines for Transparency and Openness Promotion (TOP) in Journal Policies and Practices,

or the Top Guidelines,' in short. As stated in the results and discussion sections, 22 of the 27 journals address six of the eight standards used in this study. To facilitate reproducibility, the journals sampled in this study specify key standards and guidelines in their 'Information for Authors,' or 'Guides for Authors' sections of the submission process. Because they contain all key essential elements included in the TOP Guidelines, the journals such as 1) PLoS Biology; PLoS Computational Biology, and Journal of Plant Interactions, and Cell Discovery, and Journal of Marine Science should be used as benchmark for other journals to follow suite.

As discussed above, OA biomedical sciences journals, in their manuscript submission guidelines, explicitly promote transparency and reproducibility. Some journals, for example 'Open Biology', go as far as asking authors to complete a reporting form that will be made available to editors and reviewers during manuscript assessment. The form will be published alongside all accepted manuscripts. The reporting form covers details such as information about their experiments, analyses, or data collection within the Materials and Methods and/or figure legends, to ensure that readers can easily understand what was measured and analyzed and can accurately perform the relevant protocols.

## REFERENCES

Abadal, E. (2013). Gold or green: The debate on open access policies. *International Microbiology*, *16*(3), 199–203. PMID:24568035

(2017). Announcement: Towards greater reproducibility for life-sciences research in Nature (2017). *Nature*, *546*(7656), 8. doi:10.1038/546008a PMID:28569823

Bachand, R. G., & Sawallis, P. P. (2003). Accuracy in the Identification of Scholarly and Peer-Reviewed Journals and the Peer-Review Process Across Disciplines. *The Serials Librarian*, *45*(2), 39–59. doi:10.1300/J123v45n02_04

Baker, M. (2016). Reproducibility crisis. *Nature*, *533*(26), 353–366. PMID:25993940

Bolli, R. (2017). New initiatives to improve the rigor and reproducibility of articles published in Circulation Research. *Circulation Research*, *121*(5), 472–479. doi:10.1161/CIRCRESAHA.117.311678 PMID:28819032

Bøtker, H. E., Hausenloy, D., Andreadou, I., Antonucci, S., Boengler, K., Davidson, S. M., Deshwal, S., Devaux, Y., Di Lisa, F., Di Sante, M., Efentakis, P., Femminò, S., García-Dorado, D., Giricz, Z., Ibanez, B., Iliodromitis, E., Kaludercic, N., Kleinbongard, P., Neuhäuser, M., ... Heusch, G. (2018). Practical guidelines for rigor and reproducibility in preclinical and clinical studies on cardioprotection. *Basic Research in Cardiology*, *113*(5), 1–73. doi:10.100700395-018-0696-8 PMID:30120595

Brown, P. O., Cabell, D., Chakravarti, A., Cohen, B., Delamothe, T., Eisen, M., . . . Watson, L. (2003). *Bethesda statement on open access publishing*. https://www.earlham.edu/~peters/fos/bethesda.htm

Center for Open Science. (2021). *Guidelines for Transparency and Openness Promotion (TOP) in Journal Policies and Practices "The Top Guidelines"*. Retrieved from https://osf.io/ud578/

FASEB. (2016). *Enhancing Research Reproducibility: Recommendations from the Federation of American Societies for Experimental Biology*. https://www.faseb.org/Portals/2/PDFs/opa/2016/FASEB_Enhancing%20Research%20Reproducibility.pdf

França, T. F., & Monserrat, J. M. (2018). Reproducibility crisis in science or unrealistic expectations? *EMBO Reports*, *19*(6), e46008. doi:10.15252/embr.201846008 PMID:29695587

Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). *Green and gold open access percentages and growth, by discipline.* arXiv preprint arXiv:1206.3664

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science Translational Medicine, 8*(341).

Grüning, B., Chilton, J., Köster, J., Dale, R., Soranzo, N., van den Beek, M., Goecks, J., Backofen, R., Nekrutenko, A., & Taylor, J. (2018). Practical computational reproducibility in the life sciences. *Cell Systems*, *6*(6), 631–635. doi:10.1016/j.cels.2018.03.014 PMID:29953862

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Hajjem, C., & Hilf, E. R. (2008). The access/impact problem and the green and gold roads to open access: An update. *Serials Review*, *34*(1), 36–40. doi:10.1080/00987913.2008.10765150

Harnad, S., Brody, T., Vallières, F. O., Carr, L., Hitchcock, S., Gingras, Y., ... Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, *30*(4), 310–314. doi:10.1080/00987913.2004.10764930

Justin Kitzes. (2018). *The Practice of Reproducible Research: Case Studies and Lessons From the Data-Intensive Sciences*. University of California Press.

Kannt, A., & Wieland, T. (2016). Managing risks in drug discovery: Reproducibility of published findings. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *389*(4), 353–360. doi:10.100700210-016-1216-8 PMID:26883784

Kenall, A., Edmunds, S., Goodman, L., Bal, L., Flintoft, L., Shanahan, D. R., & Shipley, T. (2015). Better reporting for better research: A checklist for reproducibility. *BMC Neuroscience*, *16*(1), 44. doi:10.118612868-015-0177-z PMID:26202681

Markowetz, F. (2015). Five selfish reasons to work reproducibly. *Genome Biology*, *16*(1), 274. doi:10.118613059-015-0850-7 PMID:26646147

Meeks, K. (2018). Ulrichsweb Review. *Journal of Electronic Resources in Medical Libraries*, *15*(1), 26–35. doi:10.1080/15424065.2018.1432436

Miyakawa, T. (2020). No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*, *13*(24), 24. Advance online publication. doi:10.118613041-020-0552-2 PMID:32079532

Morrison, H. (2017). Directory of Open Access Journals (DOAJ). *The Charleston Advisor*, *18*(3), 25–28. doi:10.5260/chara.18.3.25

Motulsky, H. J. (2014). Common misconceptions about data analysis and statistics. *Naunyn-Schmiedeberg's Archives of Pharmacology*, *387*(11), 1017–1023. doi:10.100700210-014-1037-6 PMID:25213136

Munafò, M., Nosek, B., & Bishop, D. (2017). A manifesto for reproducible science. *Nature Human Behavior, 1*, 21. doi:10.1038/s41562-016-0021

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. doi:10.17226/25303

Redalyc, L., Clase, R., & In-Com Uab, S. (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. https://www.redalyc.org/pdf/782/78241008.pdf

Rizor, S. L., & Holley, R. P. (2014). Open access goals revisited: How green and gold open access are meeting (or not) their original goals. *Journal of Scholarly Publishing*, *45*(4), 321–335. doi:10.3138/jsp.45.4.01

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. *PLoS Computational Biology*, *9*(10), e1003285. doi:10.1371/journal.pcbi.1003285 PMID:24204232

Suber, P. (2010). A very brief introduction to open access, Advancing Your Research Skills for the digital. *Age (Dordrecht, Netherlands)*, *7*. https://scholarworks.wmich.edu/digital_research_skills/7

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, *115*(3), 1231–1240. doi:10.100711192-018-2715-9

University of Cambridge, The Gurdon Institute. (n.d.). *Electronic Lab Notebooks – for prospective users*. https://www.gurdon.cam.ac.uk/institute-life/computing/elnguidance

Zhang, L., & Watson, E. M. (2017). Measuring the impact of gold and green open access. *Journal of Academic Librarianship*, *43*(4), 337–345. doi:10.1016/j.acalib.2017.06.004

## ADDITIONAL READING

Beall, J. (2012). Predatory publishers are corrupting open access. *NATNews*, *489*(7415), 179. PMID:22972258

Bradbury, A., & Plückthun, A. (2015). Reproducibility: Standardize antibodies used in research. *NATNews*, *518*(7537), 27. PMID:25652980

Dirnagl, U. (2019). Rethinking research reproducibility. *The EMBO Journal*, *38*(2), e101117. doi:10.15252/embj.2018101117 PMID:30518534

Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biology*, *13*(6), e1002165. doi:10.1371/journal.pbio.1002165 PMID:26057340

Mack, S., Cruzado-Park, I., Chapman, J., Ratnayake, C., & Vigh, G. (2009). A systematic study in CIEF: Defining and optimizing experimental parameters critical to method reproducibility and robustness. *Electrophoresis*, *30*(23), 4049–4058. doi:10.1002/elps.200800690 PMID:19960469

Piccolo, S. R., & Frampton, M. B. (2016). Tools and techniques for computational reproducibility. *GigaScience*, *5*(1), s13742–s016. doi:10.118613742-016-0135-4 PMID:27401684

Swan, A., & Brown, S. (2004). Authors and open access publishing. *Learned Publishing*, *17*(3), 219–224. doi:10.1087/095315104323159649

## KEY TERMS AND DEFINITIONS

**Computational Reproducibility:** Obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.

**Gold Open Access:** Open access delivered by journals.

**Green Open Access:** Open access delivered by repositories.

**Inferential Reproducibility:** The drawing of qualitatively similar conclusions from either an independent replication of a study or a reanalysis of the original study.

**Methods Reproducibility:** The provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated.

**Open Access:** Free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

**Reproducibility:** The ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator.

**Results Reproducibility:** Obtaining the same results from the conduct of an independent study whose procedures are as closely matched to the original experiment as possible.

**Scholarly Communication:** The system through which research and other scholarly writings are created, evaluated for quality, disseminated to the scholarly community, and preserved for future use.

**TOP Guidelines:** A suite of tools to guide implementation of better, more transparent research.

## APPENDIX

*Top peer-reviewed and referred OA biomedical sciences journals*

| | | |
|---|---|---|
| 1. AOB Plants | 10. BMC Zoology | 19. Journal of Plant Interactions |
| 2. Biology Open | 11. Cell Discovery | 20. Marine Biodiversity Records |
| 3. BioMed Central Bioinformatics | 12. EMBO Molecular Medicine | 21. Microbiology Open |
| 4. BMC Biology | 13. European Journal of Cell Biology | 22. New Disease Reports |
| 5. BMC Biotechnology | 14. Frontiers in Plant Science | 23. Nucleic Acids Research |
| 6. BMC Genomic Data | 15. Frontiers in Zoology | 24. Open Biology |
| 7. BMC Genomics | 16. Genome Biology | 25. Plant Biotechnology Journal |
| 8. BMC Molecular and Cell Biology | 17. Journal of Biological Chemistry | 26. PLoS Biology |
| 9. BMC Plant Biology | 18. Journal of Marine Sciences | 27. PLoS Computational Biology |