INVITED
doi:10.26443/seismica.v2i1.278
SEISMICA

How reproducible and reliable is geophysical research? A review of the availability and accessibility of data and software for research published in journals

Mark T. Ireland *[1], **Guillermo Algarabel**[2], **Michael J. Steventon** [3], **Marcus R. Munafò** [4]

[1]Newcastle University, School of Natural and Environmental Sciences, Drummond Building, Newcastle Upon Tyne, NE1 7RU, [2]Department of Physics, Durham University, Durham University, Stockton Road, Durham, DH1 3LE, [3]Shell Research Ltd., Shell Centre, London, United Kingdom, [4]School of Psychological Science, University of Bristol, Bristol, United Kingdom

Author contributions: *Conceptualization*: M. T. Ireland. *Conceptualization Supporting*: M. J. Steventon, M. R. Munafò. *Data Curation*: G. Algarabel, M. T. Ireland. *Formal Analysis*: G. Algarabel, M. T. Ireland. *Investigation*: M. T. Ireland, G. Algarabel, M. J. Steventon, M. R. Munafò. *Writing – original draft*: M. T. Ireland. *Writing – review & editing*: M. T. Ireland, M. J. Steventon, M. R. Munafò, G. Algarabel.

Abstract Geophysical research frequently makes use of agreed-upon methodologies, formally published software, and bespoke code to process and analyse data. The reliability and repeatability of these methods is vital in maintaining the integrity of research findings and thereby avoiding the dissemination of unreliable results. In recent years there has been increased attention on aspects of reproducibility, which includes data availability, across scientific disciplines. This review considers aspects of reproducibility of geophysical studies relating to their publication in peer reviewed journals. For 100 geophysics journals it considers the extent to which reproducibility in geophysics is the focus of published literature. For 20 geophysical journals it considers a) journal policies on the requirements for providing code, software, and data for submission; and b) the availability of data and software associated for 200 published journal articles. The findings show that: 1) between 1991 and 2021 there were 72 articles with reproducibility in the title and 417 with reliability, with an overall increase in the number of articles with reproducibility or reliability as the subject over the same period; 2) while 60% of journals have a definition of research data, only 20% of journals have a requirement for a data availability statement; and 3) despite ~86% of sampled journal articles including a data availability statement, only 54% of articles have the original data accessible via data repositories or web servers, and only 49% of articles name software used. It is suggested that despite journals and authors working towards improving the availability of data and software, frequently they are not identified, or easily accessible, therefore limiting the possibility of reproducing studies.

Production Editor:
Gareth Funning
Handling Editor:
Samantha Teplitzky
Copy & Layout Editor:
Hannah F. Mark

Reviewed by:
Daniel Nüst
Wynn Tranfield

Received:
16 November 2022
Accepted:
10 May 2023
Published:
30 May 2023

Non-technical summary In studies of the Earth, other planets, oceans and atmospheres, scientists often carry out quantitative analysis of measurements from specialist instruments or create numerical models to represent complex natural systems. These approaches are useful for understanding important processes such as plate tectonics and patterns of ocean circulation, and often have wider societal importance, such as understanding natural hazards or the distribution of economically significant natural resources. When scientists present the findings of their work in scientific publications, the focus is primarily on the written narrative. However, a cornerstone of the scientific method should be the ability to replicate an experiment or study. To enable this the input data and details of the methodology, for example the computer code used, are essential. This work reviewed how reproducible the published work in the field of geophysics has been to date. The findings show that despite most publications now requiring the underlying data to be made available, most of the time these data are not easily accessible, and therefore limit the opportunity for scientists to verify existing findings.

1 Introduction

Geophysics is perhaps best described as the application of physics to study the Earth, oceans, atmosphere, and near-Earth space, including other planets (British Geophysical Association, 2014). Geophysical methods, which typically either take raw records from instrumentation and process the recorded signals or carryout numerical modelling, rely on quantitative analysis to make robust interpretations of these systems. Frequently, geophysical methods use processing flows with numerous (often iterative) steps to accomplish tasks such as, for example, distinguishing signal from noise (Robinson and Treitel, 2000), or modelling complex processes such as mantle convection (Hager and Clayton, 1989). The reproducibility and reliability of these methods is vital to ensure that the scientific community can verify previous findings and avoid the dissemination, or misinterpretation, of results which are unreliable or ambiguous (Steventon et al., 2022). Computer analysis has long been vital to geophysical methods (cf. Reese, 1965),

*Corresponding author: mark.ireland@newcastle.ac.uk

and this continues to be true today, where most methods involve the use of code or software to process and analyze data sets of ever increasing volume.

Different scientific disciplines often use reproducibility and replicability inconsistently (National Academies of Sciences, 2019). In geophysics, and Earth Sciences more broadly, definitions and terminology used for reproducibility, replicability and reliability in research have not been examined to the same extent that they have been in, for example, medical sciences (e.g., Goodman et al., 2016). The Turing Way project defines reproducible research as "work that can be independently recreated from the same data and the same code that the original team used" (Arnold et al., 2019). It is useful to expand this definition by classifying how reproducible research is different from robust, replicable, and generalisable research (Figure 1). The Turing Way definitions for each of are as follows:

- **Reproducible**: when the same analysis steps performed on the same dataset consistently produces the same answer.

- **Replicable**: when the same analysis performed on different datasets produces qualitatively similar answers.

- **Robust**: when the same dataset is subjected to different analysis workflows to answer the same research question and a qualitatively similar or identical answer is produced. Robust results show that the work is not dependent on the specifics of the programming language/equipment/methodology chosen to perform the analysis.

- **Generalisable**: Combining replicable and robust findings allow us to form generalisable results. Generalisation is an important step towards understanding that the result is not dependent on a particular dataset nor a particular version of the analysis pipeline.

To date the existing published literature on the topic of reproducibility in geophysics can broadly be grouped into four areas: 1) the benefits of specific open-source software for improved repeatability (e.g. Oren and Nowack, 2018); 2) the repeatability of surveying techniques (e.g. Waage et al., 2018); 3) the reproducibility of individual studies (e.g. Walker et al., 2021); and 4) improving the repeatability of specific workflows (e.g. Jun and Cho, 2022). There has been some limited examination of reproducibility in geosciences more broadly (e.g., Konkol et al., 2019; Nüst and Pebesma, 2021; Steventon et al., 2022). Specifically in the field of geophysics, however, to date there has been no empirical consideration of the extent to which the existing publications and published work are reproducible.

The role of journals has been described as a way to provide a source of information and knowledge that can easily be located and read (Childe, 2006). This includes specifically: 1) registration of the author's claim to the work; 2) certification, usually by peer review, that the research was conducted properly; 3) dissemination; and 4) archiving, providing a permanent public record of the work that can be found and cited (Rallison, 2015). Preserving the data and methods that underpin research has become an increasingly important part of the publication process. In some subjects journals have acknowledged the need to strengthen approaches to reproducibility (Nature, 2014) and some even adopt specific policies with regards to verification of results (e.g., American Journal of Political Science, 2019). Similarly, the Transparency and Openness Promotion (TOP) guidelines were introduced by the Centre for Open Science to review the extent to which the research that journals support improves reproducibility through increasing the transparency of the research process (Nosek et al., 2015).

This work attempts to quantify the ways in which existing journals that publish geophysical research have to date made data and software available and accessible, in turn promoting reproducibility and repeatability.

2 Reproducibility, repeatability, and data availability

In recent years there has been increased attention on aspects of reproducibility, including data availability, across many scientific disciplines (e.g. Tedersoo et al., 2021); however, there has been limited focus on this topic in areas of Earth Sciences (Wildman and Lewis, 2022). At the 2016 G20 Summit, the G20 leaders formally endorsed the application of FAIR principles to research data (European Commission, 2016). The FAIR principles set out the importance of research data being Findable, Accessible, Interoperable and Reusable to improve and accelerate scientific research (Hodson et al., 2018) and were set out by a diverse set of stakeholders across academia, industry, funding agencies, and scholarly publishers. Contemporaneous to this, computational approaches have become increasingly important as more and more scientists are now able to adopt computational methods due to the improved ease and availability of both hardware and software (cf. Mesirov, 2010). Indeed, software is now a ubiquitous, if often invisible, component of research in most scientific disciplines, and for research to be reproducible requires understanding the software used by the original research (e.g., National Academies of Sciences, 2016). With this, the availability and support for large scale data sharing has led to increased attention and resources to enable scientists to share data (Tenopir et al., 2011). Despite computational and storage infrastructure being in place, there are still perceived barriers to effectively making both data (Tenopir et al., 2011) and software (Gomes et al., 2022) available and accessible. In a survey of >1300 scientists on data sharing practices, Tenopir et al. (2011) found that one third of the respondents chose not to answer whether they make their data available to others, and of those that did respond 46% reported they do not make their data electronically available to others. In exploring why researchers chose not to make their data available Tenopir et al. (2011) found

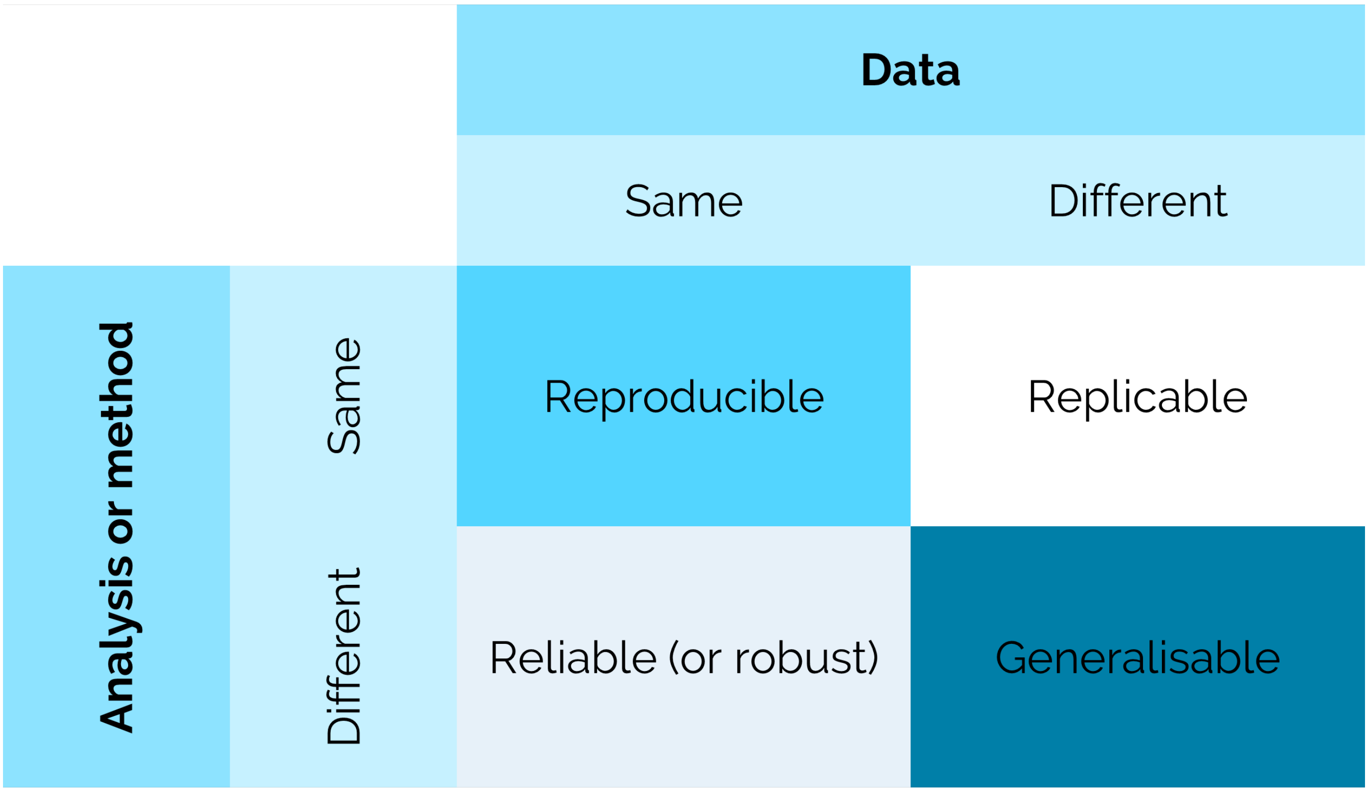

Figure 1 Definitions of reproducible, replicable, reliable, and generalisable, mapped against data and methods. Modified after Arnold et al. (2019).

the leading reason is insufficient time (54%), followed by lack of funding (40%), having no place to put the data (24%), lack of standards (20%), and "sponsor does not require" (17%), with only 14% of respondents stating their data "should not be available". For code sharing, Gomes et al. (2022) identified reasons why code sharing is not more common in biological sciences, including perceived barriers such as: unclear process, complex workflows, data too large, lack of incentives, and concerns on re-use of data. These barriers are also identified in other disciplines, for example psychology (Houtkoop et al., 2018), and more broadly across the science community (Borgman, 2010).

3 Review Methodology

This study considers the reproducibility of geophysical studies which have been published in peer reviewed journals. It does not include any consideration of the reproducibility of geophysical studies outside of this, for example unpublished work from the private sector, or non-peer reviewed published reports. The analysis consists of three parts: 1) a mapping review of the extent to which reproducibility in geophysics is explored in the literature; 2) a review of journals' policies on the requirements for providing code, software, and data for submission; and 3) a sample of articles examining the availability of code, software, and data.

Each part of the analysis is based on geophysical journals as identified by SCImago Journal Rank (see Supplemental Table 1 in Ireland, 2022). SCImago Journal Rank (SJR) is a numeric value representing the average number of weighted citations received during a selected year per document published in that journal during the previous three years, as indexed by Scopus (SCImago). While journal metrics are frequently misused to assess the influence of individual papers (Pendlebury, 2009), here the list is simply used as a mechanism to firstly identify journals by subject area. Each journal in the list is assigned a subject area and subject category. We include journals where either the first or second subject category is "geophysics". The journals identified using SCImago are a broad representation of journals which may be widely read and used by the geophysics community, or they frequently publish articles where geophysics is the dominant discipline. Journals whose exclusive focus are review articles are excluded from the analysis. The review does not use the SJR as a measure of the 'prestige' of any individual journal, nor to make any comparison or interpretations between individual journals.

3.1 Mapping Review Protocol

This study used a basic mapping review, designed to identify primary studies relating reproducibility and reliability in geophysics without manually selecting which articles to include. The aim was to enable a semi-quantified assessment of the extent to which studies focus on the topic of reproducibility (or reliability) in geophysics and determine how frequently the primary focus of studies is to investigate reproducibility or reliability. To do this, search strings were constructed based on the terms "reproducibility", "reliability" and "replicable". The search strings used are as shown in Table 1.

We restricted the search to the journals ranked in the

Definition theme	Search terms
Reproducibility	reproducibility OR reproduce OR reproducible OR reproduction
Reliability	reliability OR reliable OR reliably OR reliabilities

Table 1 Search strings used in the literature mapping review.

top 100 by SCImago (see Supplemental Table 1 in Ireland, 2022). The searches are conducted using Publish or Perish software (Harzing, 2010, 2007), using a single search for each journal. The searches used Google Scholar and while there is still no consensus on the use of Google Scholar in systematic literatures reviews (Boeker et al., 2013), it is adopted here as it is free-to-use, and therefore allows anyone to repeat the searches carried out in the future, regardless of access to subscription services. For each search the date range and title of the journal was specified. The terms in Table 1 were used for title word searches only. The data presented are accurate as of 3 April 2023.

3.2 Review of Journal Policies

To evaluate journals' existing policies relating to the inclusion of code, software, and data, we use the top 20 entries on the list of geophysical journals identified using SCImago Journal Rank. For each of these journals the requirements for code, software, and data, as per the 'instructions for authors' and the publishers' policies, were compiled. Table 2 shows the criteria for which we reviewed journals' policies. As the criteria outlined in Table 2 are rarely a clear binary yes/no, scoring criteria were used. The scoring criteria used are shown in Table 3. The score for each journal was assessed by author Algarabel and then reviewed by author Ireland. It is acknowledged that using scoring criteria like this could be considered subjective; however, by using descriptors of the criteria it is anticipated that aspects of bias are minimized.

3.3 Review of Journal Submissions

To evaluate the extent to which published, peer-reviewed articles make data and code available, the same list of the top 20 geophysical journals identified using SCImago Journal Rank was used (see Supplemental Table 1 in Ireland, 2022). As journals do not currently include search filters to discern which articles make data available, a sample of individual publications was selected to evaluate the extent to which they meet the criteria set out by a journal's policy. Two hundred articles were selected from a 3-year period (2020-2022). Again Publish or Perish software (Harzing, 2010, 2007) was used, and Google Scholar used as for the search. The date range was set to 2020-2022[1] and the "Maximum number of results" was set to 10. This search was carried out for each of the 20 journals in Supplemental Data Table 1 (Ireland, 2022)[2].

Each article is noted as either open access or paywalled. This is on a per article basis, rather than by journal, since authors may opt to make an article in a subscription access journal available open access by paying a journal an Article Publication Charge (APC). Again, as the availability and accessibility rarely be described using binary yes/no, scoring criteria are used, shown in Table 4. To assess the availability and accessibility of software we used the same sample of 200 articles as for data and reviewed if an article named any software used in the research. We searched the main text, availability statements, acknowledgements, and supplementary materials (where present). We also, where possible, report the license of the software that was used (e.g. open source or commercial). Throughout the article the word software is used as an inclusive term covering applications with graphical user interfaces (GUIs), code for interpreted programming languages (e.g. Python), and code for general-purpose programming languages (C++).

Included in policy/guidance	Category
Has definition of 'research data'	Policy
Includes separate 'data policy' section	Policy
Requirement to include data availability statement	Data
Requirement to include citations for data	Data
Requirement to make data available	Data
Guidance to include data in dedicated data repository	Data
Requirement to include software/code availability statement	Software
Requirement to include citations for software/code	Software
Requirement to make software/code available	Software
Guidance to include software in dedicated repository	Software
Guidance to include data in supplementary materials	Data

Table 2 Criteria which journal policies and guidelines were reviewed against.

We made the decision to anonymise the data as we considered that this review was centered on the field of geophysics rather than highlighting the reproducibility of any individual published piece of work. Although the articles were anonymised we maintained a key to enable us to link the anonymised list back to the original sources.

In the 100 journals which publish geophysical research searched there were, between 1991 and 2022, 72 articles with "reproducibility", "reproduce", "reproducible" or "reproduction" in the title and 417 with "reliability", "reliable", "reliably", or "reliabilities" in the title (see Figure 2). From 1990 to 1999 there were 64 publications with "reliability", "reliable", "reliably", or "re-

[1]The search was done on 29 July 2022, and therefore covers articles published and index over 31 months.

[2]Despite having introduced those dates, some articles date from 2019.

Score	Summary	Descriptions
1	Required	Required, (e.g., must) with very limited exceptions (for example to preserve confidentiality of human participants)
2	Partial requirement	Partial requirement with flexibility around inclusion method.
3	Encouraged	Encouraged, with wording proactively encouraging (e.g., *should*) authors to include
4	Mentioned	Mentioned or implied but not proactively encouraged
5	Not mentioned	No mention in guidance to authors
6	Not allowed	Inclusion of data or content not permitted.

Table 3 Scoring criteria used to evaluate the extent to which journals proactively support improving the availability of data and code.

Score	Summary	Descriptions
1	Data available and accessible via dedicated data repository	Data available and is hosted on a repository which provides a DOI for the data. Includes where data is provided in tables within article.
2	Data available via website / webserver	Data available but no DOI.
3	Data source linked	Includes cases where article provides link to a web-hosted database but the specifics of the dataset (for example time periods, filters) are not clear.
4	Data provided in supplementary information or data	Includes where data are included under 'supplementary information'. The lack of consistency in use of supplementary information makes data frequently harder to access.
5	Data listed as available but not accessible	Includes when authors state 'data available on request'.
6	Data not available or no mention of data availability	Includes when authors explicitly state that data is confidential and not available or accessible.
X	Data linked but link no longer valid	

Table 4 Scoring criteria used to evaluate the availability and accessibility of data in published articles.

liabilities" in the title. Compare this with 2000 to 2009, when there were 114, and 2010 to 2019 when there were 181. This represents an increase of 77% and 59% respectively. From 1990 to 1999 there were 8 publications with "reproducibility", "reproduce", "reproducible", or "reproduction" in the title, between 2000 to 2009 there were 13 and between 2010 to 2019 there were 34. These represent an increase of 63% and 161% respectively.

Of the 100 journals, 32 (32%) have published articles with "reproducibility", "reproduce", "reproducible", or "reproduction" in the title, and 64 (64%) have published articles with "reliability", "reliable", "reliably", or "reliabilities" in the title. The Bulletin of Earthquake Engineering has published the most articles with "reliability" "reliable", "reliably", or "reliabilities" in the title (63). Geophysical Research Letters has published the most articles with "reproducibility", "reproduce", "reproducible", or "reproduction" in the title (11). A full breakdown of the number of publications with keywords in the title is provided in Supplementary Data Tables 2 and 3 (Ireland, 2022).

3.4 Journal Policies

From reviewing journal policies, it was found that 12 out of 20 (60%) journals have a definition of research data, while eight out of 20 (40%) do not have a definition (see Figure 4); 17 out of 20 (85%) of journals have a discrete 'data' section within the journal policies and guidance. Despite 18 of the 20 journals either requiring or mentioning making data available, eight of these are from a single publisher, the American Geophysical Union (AGU), which applies the same requirements across all its Earth science publications. Only four out of 20 (20%) have a requirement for a data availability statement and only one journal, The Journal of Petrology, has an explicit requirement for both inclusion of data and a data availability statement. Information for authors is found within dedicated data policy sections for 17 out of 20 (85%), with three (15%) embedding the information within other sections. These results are summarized in Figure 5.

It is found that only one of the 20 journals (5%) reviewed required any code used to be made available and only one out of 20 journals (5%) require a code availability statement. There are 12 out of 20 (60%) journals that encourage making code available, while seven out of 20 make no mention of making code available. No journals have a requirement to make data or code available through repositories, or to include DOIs. However, 15 of the 20 journals (75%) encourage the use of data repositories and 14 of the 20 journals (70%) encourage the use of DOIs. Two of the 20 journals (10%) mention the use of repositories, and four of the 20 (20%) mention the use of DOIs. Two of the journals (10%) make no mention of the use of repositories and one journal makes no mention

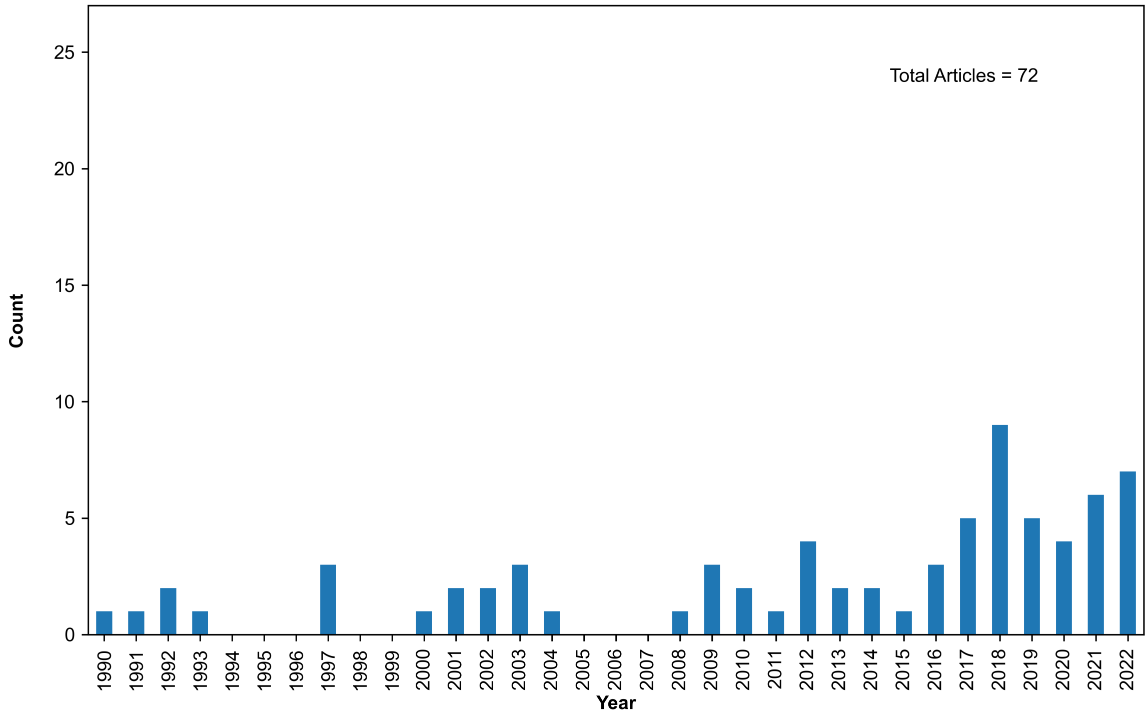

Figure 2 Number of publications with the keywords “reproducibility”, “reproduce”, “reproducible” or “reproduction” in the article title, by year.

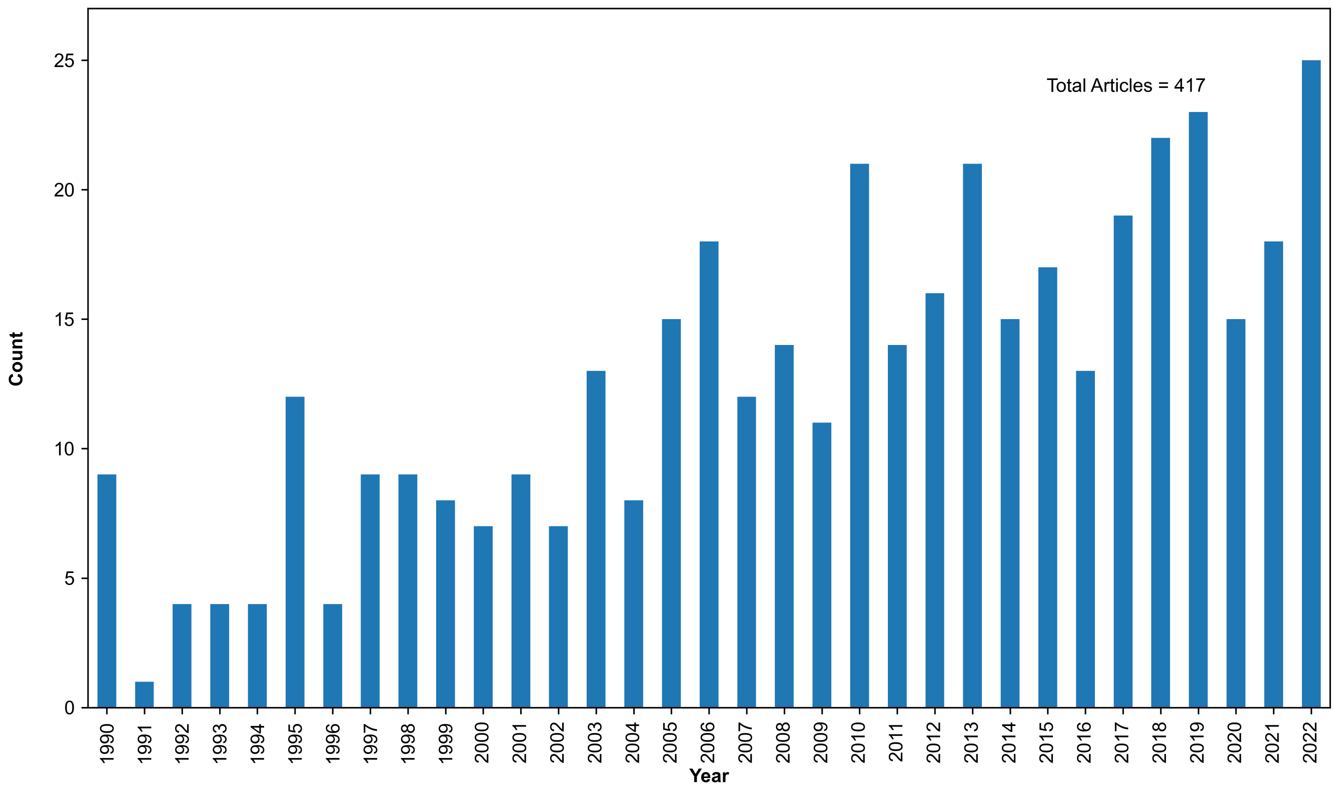

Figure 3 Number of publications with the keywords “reliability”, “reliable”, “reliably” or “reliabilities in the article title, by year.

of the use of DOIs.

Qualitative analysis of journals policies and guidance suggests that different publishers are adopting different approaches to encouraging making data and code available. Some are clear that they now require the inclusion of available data. For example, the AGU author resources explicitly refer to the FAIR principles and include the following regarding data availability statements:

"It is not sufficient to write that your data will be available upon request and to archive and make your data available in the supplementary information of your manuscript." (AGU)

In contrast, the Society of Exploration Geophysicists (SEG)'s Geophysics makes no direct reference in the author instructions to the FAIR principles, although the SEG is a signatory to the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) Statement of Commitment. In their instructions to authors, they state:

"... papers from industry authors and academic researchers whose work is built on unshareable industry-owned data are invited, encouraged, and welcome." (Geophysics)

The guidance for authors across journals frequently allows for authors to self-select from a range of options relating to data availability; however, only in the case of two publishers, AGU and Springer, was there any text indicating that the deposition of data was checked as part of the publishing process.

3.5 Journal Submissions

Of the articles with accessible information, it is identified that 165 of the 191 (~86%) articles have data availability statements and 26 (~14%) do not have data availability statements. A breakdown of data availability statements by journal is shown in Table 5. All sampled articles (n=100) published across the 10 AGU journals included data availability statements. In contrast, of the eight sampled articles accessible to us published in Economic Geology, only one had a data availability statements, and seven had no data availability statement.

Of the 191 articles sampled, 90 (~47%) make available original data from their research and a further nine (~4%) provide information to available secondary data sources. Four articles state that the original data is available on request and four articles state that secondary data is available on request. Four articles provide no information of the availability of original data, and 38 articles provide no information on the availability of secondary data. 41 of the 191 (~21%) articles have the data available via repositories and 63 of the 191 articles provide weblinks to data sources. Zenodo, FigShare and Mendeley are the most used repositories for data sharing (~75%). Examples of data sources for which articles provide weblinks to include NASA's Planetary Data System, Incorporated Research Institutions for Seismology, and the National Oceanic and Atmospheric Administration data portal. In most instances the exact details of the dataset or search criteria used to return a dataset are not included. For articles sampled from Geophysics, Marine and Petroleum Geology, and Economic Geology, none of the articles reviewed had made the original data accessible or available.

Of the 200 articles, 132 were open-access (e.g., accessible through the publishers' site without subscription access) and 68 were paywalled access (e.g., required a subscription to access the full article). Of the 132 open access articles it was found that 46% made the data available (scores 1 to 4 in Table 3). and 54% did not make the data available via a data repository (scores 5 and 6 in Table 3). Of the 68 paywalled articles, we found that just 14% of these made their data available via a data repository or web server (Figure 6).

There is, at least qualitatively, a difference in the data availability between geophysical research which has a basis in resource or economic applications, and those with either a fundamental or global seismological focus. For example in SEG's Geophysics, which publishes research focused on geophysical method applied to extractive or resource industries (Geophysics), it was found that none of the ten articles reviewed made the underlying data available. In contrast in the Seismological Society of America's Seismological Research Letters, whose scope covers topics of broad interest across seismology and related disciplines, it was found that seven of the ten provided links to underlying data, and the three which did not, did not use original data. It is also found that for paywalled articles, publishers take different approaches as to what information to provide in the public domain. For example, in both Tectonophysics and Earth and Planetary Science Letters published by Elsevier, in some instances the data availability statement is not behind the paywall even if the full article is, whereas Geophysics, published by SEG, does not make this information available without paid access to the article.

Of the 200 articles it was found that 49% (98 articles) named software used in the research and 30% (60 articles) did not name any software used in the research (Figure 7). Of the 200 articles, we were unable to review the software used for 13.5% (27) of them due to articles being paywalled. For 6% (12) software could be considered not applicable due to articles being review papers and 1.5% (3) used large scale numerical models, where it was not possible to identify the software environment. Of the 98 articles which did name the software, 63% exclusively or partly used open source software and 38% exclusively or partly used commercial software (these do not total 100% due to some publications using a combination of open source and commercial software) (Figure 8). There were 100 unique software items identified in the 98 articles that named the software used. Of the software named those with more than five occurrences were: Python (17), Matlab (8), Generic Mapping Tools (7) and ImageJ (5).

4 Discussions

The identification of 489 articles that examine aspects of reproducibility and reliability since 1990 qualitatively suggests that both are topics of interest for geophysics research. It is worth noting that, as has long been recognized (e.g., Carr et al., 1997), the digitization of journal formats has resulted in an increase in the number of publications. Therefore, the observed increasing number of articles examining reproducibility and reliabil-

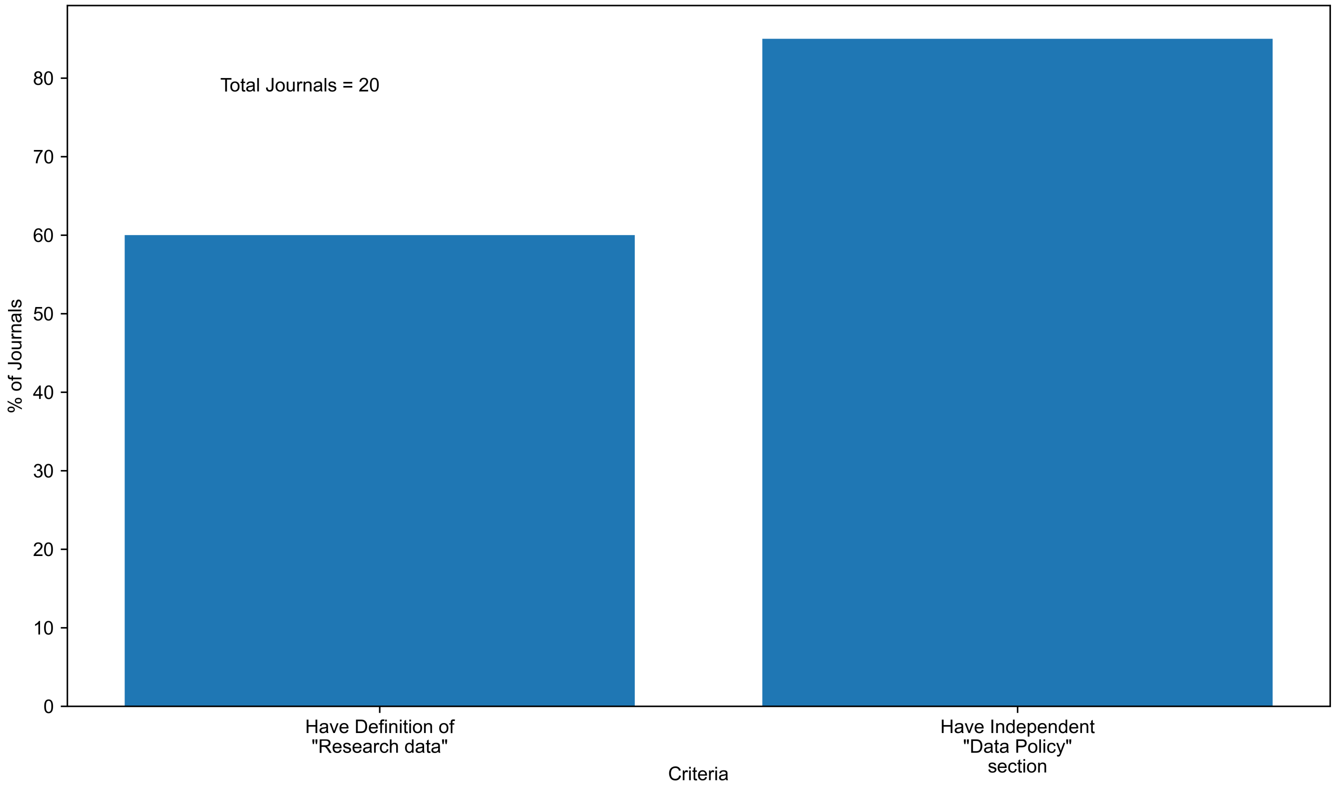

Figure 4 Percentage of journals which have a definition of "research data" and percentage of journals which have an independent "data policy" section.

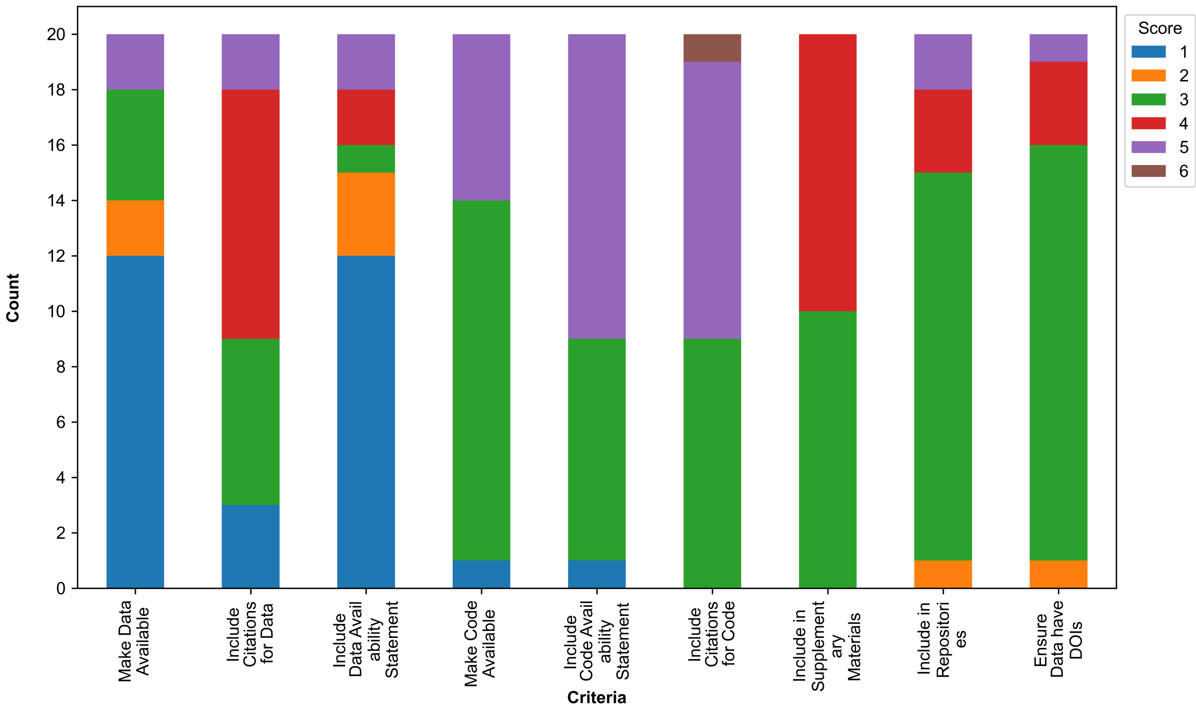

Figure 5 Stacked bar charts showing the requirements set out by journals relating to data and software. All based on the data provided in Data Table 4.

ity may not only be the result of increased attention by researchers. Of the 100 journals reviewed 32 (32%) have published articles with "reproducibility", "reproduce", "reproducible", or "reproduction" in the title, and 64 (64%) with "reliability", "reliable", "reliably", or "reliabilities" in the title, suggesting that the theme of reliability has been of greater focus than reproducibility. This would seem to support the hypothesis of Steventon et al. (2022) who suggested that studies focused on reproducibility or replication are less likely to be

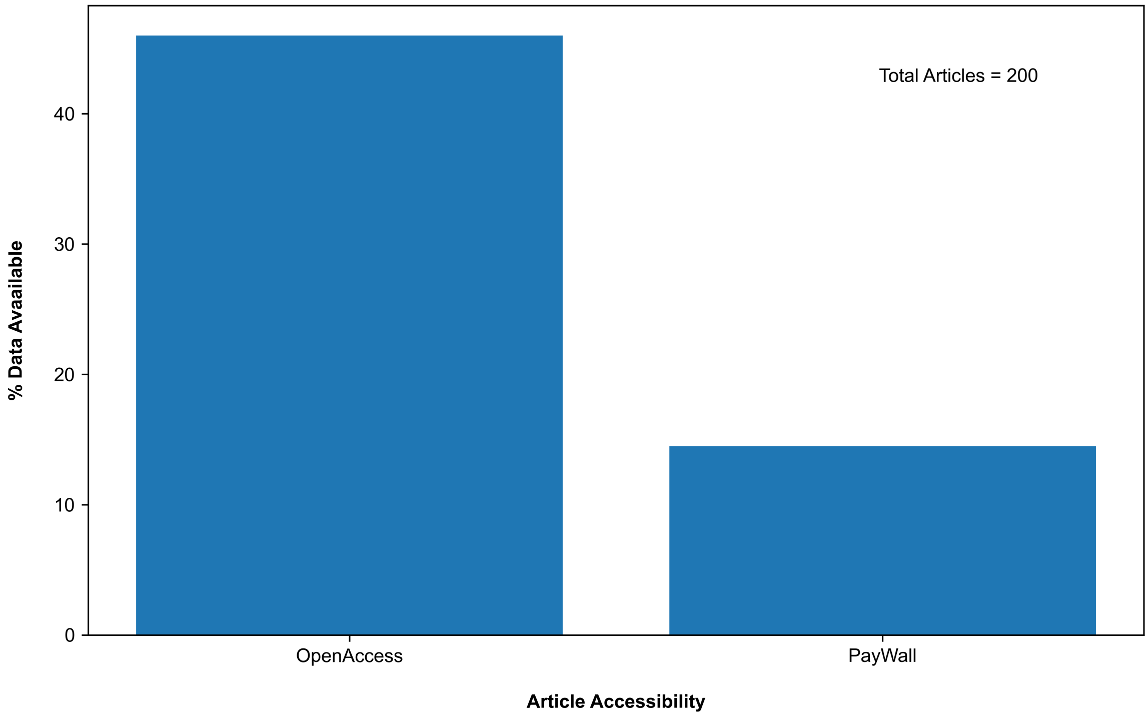

Figure 6 Chart showing the difference in data availability between open access articles and paywalled articles.

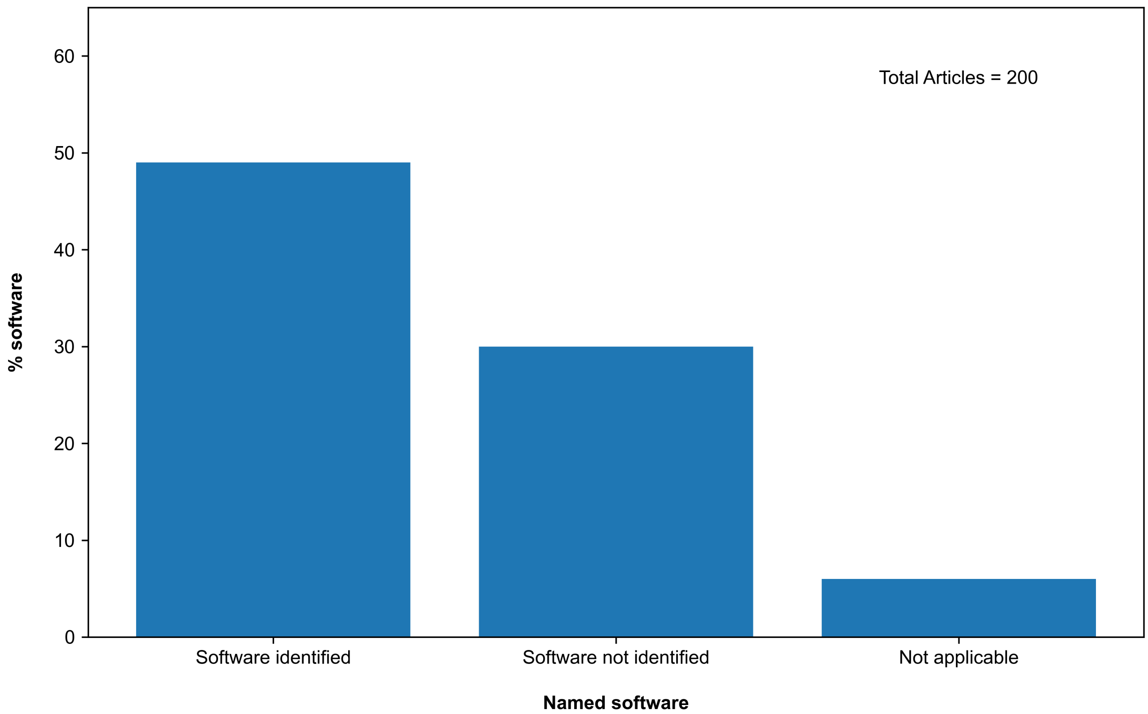

Figure 7 Chart showing the percentage of total number of articles that either identify or do not identify the software used in the research, or where the software is potentially not applicable (e.g. review articles).

published than "novel" or "ground-breaking" work. It therefore may be that this publication bias has led to published articles focusing on new methods and new datasets, rather than exploring the reproducibility and replicability of previously published research.

Journal	Data Availability Statement		Original Data Accessible	
	Yes	No	Yes	No
Tectonics (10)	10	0	8	2
Geochemistry, Geophysics, Geosystems (10)	10	0	10	0
Geophysical Research Letters (10)	10	0	2	0
Journal of Geophysical Research D: Atmosphere (10)	10	0	10	0
Journal of Geophysical Research B: Solid Earth (10)	10	0	9	0
Journal of Geophysical Research E: Planets (10)	10	0	6	0
Journal of Geophysical Research C: Oceans (10)	10	0	1	1
Journal of Geophysical Research F: Earth Surface (10)	10	0	4	2
Earth and Planetary Science Letters (10)	10	0	10	0
Tectonophysics (10)	5	5	5	4
Geophysics (8)	7	3	0	9
Journal of Petrology (10)	10	0	9	0
Seismological Research Letters (5)	7	3	7	0
Contributions to Mineralogy and Petrology (10)	10	0	10	0
Journal of Geodesy (10)	8	2	2	3
Mineralium Deposita (10)	9	1	8	1
Economic Geology (8)	1	7	3	5
Earthquake Spectra (10)	5	5	5	2
Marine and Petroleum Geology (10)	8	2	1	5
Geophysics Journal International (10)	7	3	3	5

Table 5 Summary data for articles examined, showing the number of articles that 1) provided a data availability statement and 2) whether they made the original data available. For both criteria scores 1 to 4 count as 'yes' and scores 5 and 6 counted as 'no' (see Table 3 for details on scoring). As not all articles used original data, or some were solely modelling studies, the total of yes/no for *original data* does not always match the total count.

4.1 Subjective interpretation of journal policies

The findings indicate that journals have a mixed approach to the wording used in policies relating to the provision of data and code (see Figure 5). While the TOP Factor (https://topfactor.org/) provides a score of how journal policies align scientific ideals with practices, no geophysical journals have been scored to date (as of 17 April, 2023). We found that journals repeatedly used ambiguous language in their policies when referring to data and code availability. While 60% of journals had a policy which stated that the submission of data was a requirement, the statements used in the other 40% of journals were frequently ambiguous, using terms such as encourages, where possible, where applicable. Clearly those journals without a clear definition of data will likely result in more subjective interpretation of the guidelines by authors, reviewers, and editors. From the publisher's side, from a marketing and commercial perspective it could make sense to have submission guidelines and policies that clearly define data and code access. A counter view could be that ambiguity in the policies and guidelines may be beneficial from a commercial perspective as it may encourage submissions and consequently facilitate the journal to publish more articles than if there were tighter restrictions on data and software requirements. However, where data and code are easily identifiable and accessible, there is empirical evidence to suggest that the sharing of research data may can be associated with an increase in citations (Christensen et al., 2019; Piwowar et al., 2007). When it comes to the use of supplementary information, it is worth highlighting, as in the AGU's data availability statement, that this section of a manuscript is still indicated as a suitable place to accommodate data. There are however issues with this as highlighted by previous studies (Pop and Salzberg, 2015). Most notably there is often a lack of guidance on how supplementary information should be used to include data (e.g. Pop and Salzberg, 2015), which means that often data or metadata provided in supplementary information is inaccessible. Supplementary files are for the most part not considered to be a part of the formal record of an article, and therefore the integrity of these materials is frequently poorly maintained. In the case of internet hosted materials, this is evidenced by other studies (e.g., Evangelou et al., 2005) which have shown that the percentage of inactive links to supplementary informa-

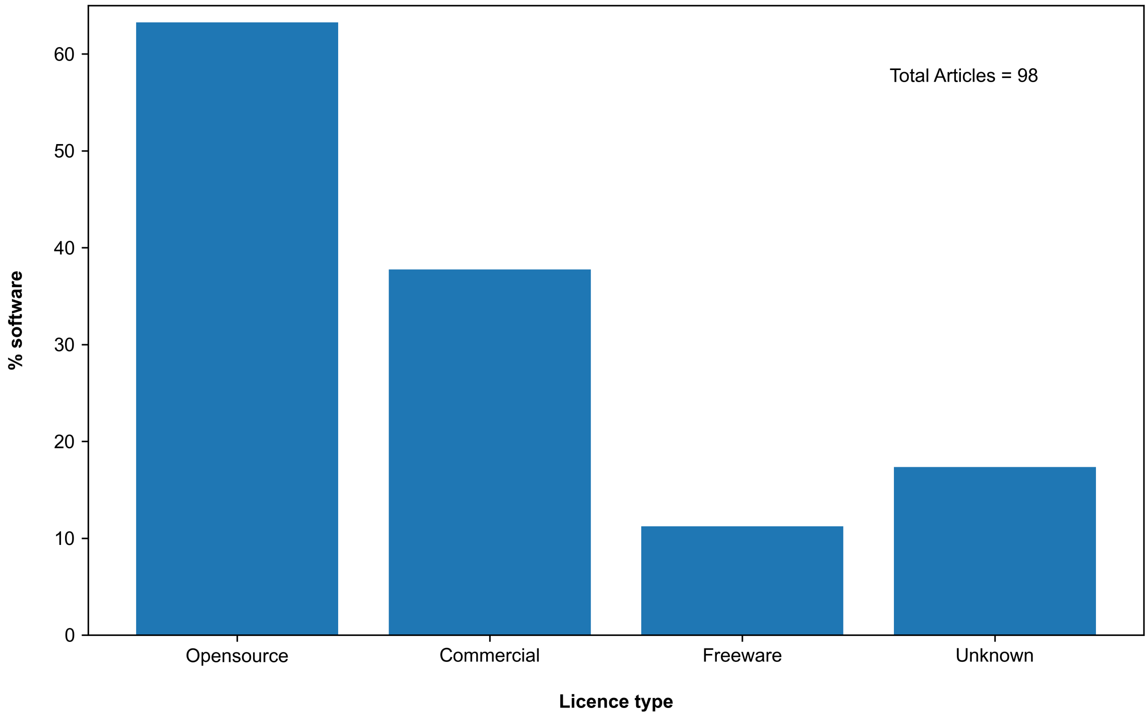

Figure 8 Chart showing, for the articles which did identify software, the percentage that make use of open source, commercial, and freeware, as well as the percent where the license was not easily identifiable or traceable. The percentages here do not total 100% as some articles used multiple pieces of software.

tion increased with time since publication,

4.2 Availability vs accessibility

We found that, where journal articles used original data, in general the availability of data was improved over journal articles which used existing data or data derived from third party sources. Frequently where articles used non-original data, while articles provided information on the data in the data availability statements, they provided insufficient information to identify specific datasets, or in several cases the weblinks no longer worked. This suggests that it is not only data availability that is important, but also data accessibility. Starr et al. (2015) list eight core principles of data citation which have been endorsed by 87 scholarly societies, publishers and other institutions. Of relevance to the findings here are the unique identification and specificity and verifiability. For the majority of the articles sampled, there was insufficient information for the dataset to be identified without human search input, in contrast to the recommendation that data identification should be machine actionable (Starr et al., 2015). Commonly it was difficult to identify the specific dataset used in the research: for example, it was possible to follow a weblink to website which hosts data, but not to identify the data on which the analysis was based. While many articles (>30%) provide weblinks or the names of the organizations which host the data, they frequently provided insufficient information for readers to identify and verify that the data is the same as that which was used by the authors. Frequent issues include, for example, the data linked consisting of multiple files with no explicit statement of what files from that dataset the authors used. Another persistent issue is the use of non-static weblinks for data sets.

4.3 Role of Journals, Editors and Reviewers

The contributions of editors and reviewers for journals, whether they are for-profit or not-for-profit, are invaluable in ensuring the continued and timely publication of scientific findings. In most cases, those scientists that undertake the role do so without remuneration. The role of a journal editor could be summarized as to sustain integrity in published research and enforce the policies and the standards for the journal, both for authors and reviewers (Caelleigh, 1993). The role of reviewers could be summarized as ”evaluat[ing] whether or not there is a meaningful...contribution, whether the constructs are adequately defined,...and whether the underlying mechanisms/process...are explained” (Lepak, 2009). Based on journal (and publisher) policies it is unclear as to whether reviewers are expected to evaluate the suitability of data and software availability statements. It could be suggested that there should be a clear distinction then between the role which editors and reviewers have in determining whether an article’s approach to data and software availability (and accessibility) is suitable or not. In practice, clarification by journals over the role of reviewers and editors could improve the situation. For example, one possibility could be for reviewers to have the responsibility for ensuring that the data and software is suitable to demonstrate the scientific findings, and that the editorial board and office has the responsibility to ensure that authors have included a data and software availability statement and adhered to the requirements for making data and software accessible and available. Indeed, this is how AGU handles the availability of data, as indicated on their information to authors where it clearly states, "AGU now checks to see if data/software has been properly cited vs simply linking to a DOI, website, platform" (AGU).

It is worth noting that while it was found that fewer journals had dedicated requirements for software, sometimes they are mentioned within the policies, guidelines, and definitions of data. This can lead to some ambiguity when the guidance is interpreted by authors. Further, while not all studies use bespoke software (e.g. customized code), there are very few aspects of geophysical research which do not have some reliance on computer-based analysis. Therefore, journals could perhaps consider a simplified approach when it comes to more commonly used software (e.g., for statistical analysis), whereby authors simply choose from a list.

In the review of existing journal submissions, it became clear that it is currently not possible to identify which articles have accessible data and software quickly and efficiently. For the most part, journals use data availability statements, with only 10% of the 20 journals examined not mentioning including a data availability statement. However, it is not possible to filter or search articles by the information in these statements. In chemistry it has been suggested that one solution to this challenge would be to completely recast data-rich scientific journal articles into two components, a narrative and separate data component, each of which is assigned a persistent digital object identifier (Harvey et al., 2014). However, perhaps a simpler solution could be the requirement for authors to choose from pre-defined categories of data availability, with authors' assertions checked for accuracy as part of the editorial process. Journals could then implement a search criterion based upon if the data is available and accessible.

4.4 Software Availability and Accessibility

In the review of existing journal submissions approximately half the articles reported the software used in the research (49%). Where identified, the software is not consistently reported in the same location in different journals, or even within different articles in the same journal. For example, some articles reported software in the 'methods' section of the article, others referenced the software used in the acknowledgements, and some only mentioned the software within supplementary materials. The 51% of articles which did not report the software used all frequently included quantitative or statistical analysis, and while articles commonly detail the theory, they do not report on the implementation of this. In other science disciplines, studies have highlighted the need for consistently specifying the analytical software used in, as different software packages could produce varying results (Dembe et al., 2011). It is postulated that where software, both commercial and open source, are widely available, accessible, and used, such as Microsoft Excel, authors may unintentionally omit them from inclusion from the methods. However, the accuracy of statistical methods in such packages has been the focus of repeated studies (e.g., McCullough and Heiser, 2008; McCullough and Wilson, 2002, 2005; Mélard, 2014). The data indicate that open source and freeware software, sometimes referred to as free and open-source software (FOSS) was used in 63% of the articles which identified the software they used. While there has been a widespread adoption of FOSS documented (e.g., Glynn et al., 2005; Hauge et al., 2010; van Rooij, 2011), there has been very limited focus on the extent to which FOSS is adopted with Earth sciences and geophysics specifically. The findings in this study suggest that commercial software still is important within in research, where 38% of the articles which identified their software made use of it. Some authors (e.g., de Groot and Bril, 2005) have speculated that FOSS has rarely been used for larger scale, high end-user applications and software is frequently closed source or proprietary. However, increasingly open source interpreted programming languages such as Python through the wide variety of packages available are increasingly capable of handling large and complex datasets such as N-dimensional arrays (e.g., Hamman, 2017). Anecdotally, the number of downloads for dedicated geophysics Python packages, suggests that open-source software is growing in usage. For example Obspy (Beyreuther et al., 2010) has, according to PePy (https://pepy.tech/) been downloaded 1,783,753 (as of 8 April 2023). Proprietary software may offer benefits, such as well-developed GUIs that do not require as high a level of computer literacy (e.g., Muenchow et al., 2019). As noted by Nüst and Pebesma (2021), in some instances, such as where software are linked to hardware, proprietary software may be unavoidable. This could include, for example, software linked to specific seismic acquisition systems.

4.5 Perceived Barriers

Data and code sharing are often perceived as being limited by digital infrastructure (Gomes et al., 2022). However, while making data and code available may have been previously limited by such restrictions, there now exits the underlying digital architecture to, for example, host individual files typically up to 20GB in size on data repositories such as Figshare and Zenodo. Repositories have added the functionality to archive code, for example from GitHub to Zenodo, and assign a DOI. Indeed, many of the perceived barriers, for example challenges in handling large data files, are not unique to geophysics and these concerns have mostly been shown to be relatively straightforward to manage in terms of absolute volume. For example, a study in neurosciences by Poldrack and Gorgolewski (2014) described how the sharing of raw MRI data from 1,000 authors would consist of ~2.7 terabytes, a relative modest volume by current storage infrastructure solutions (e.g., Behnke et al., 2019); however, there are major challenges in ensuring that data sets are curated to make them accessible and useful to researchers. Indeed, the common occurrence of big data within nearly all subjects has served to identify that discussing absolute data volume as a barrier in any context is limiting, as computing hardware and software advances at such a rate that any absolute numbers are soon superseded (Oguntimilehin and Ademola, 2014).

4.6 Limitations of study

The findings presented in this review are not exhaustive. There exist several limitations to the study that should be highlighted. Firstly, there are alternative ways in which the choice of journals to include could be made. The approach here, as far as possible, was designed to avoid user bias in the selection of journals, but it is recognized that the breadth of journals included covers some topics that may be considered outside of the immediate subject area of geophysics. Secondly, and related to this, the choice of search tools could impact the results. In this study searches were undertaken using tools and databases which did not require paid subscription access. Alternative subscription-only search services may result in different results, for the review of existing literature. Thirdly, when reviewing journal policies, there is a component of subjectivity in the categorization of a journal's requirements. As discussed above this is itself is one of the issues which publishers and journals need to tackle to avoid any ambiguity in the requirements. Fourthly, when categorizing the availability of data for an individual article, while in some cases it is very clear if data is available and accessible (e.g., DOI linked data) or not (e.g., data is confidential) there are examples where, for example the availability of the data is insufficiently described to easily assess if the data is accessible. Examples of this include where a link to a website which hosts data is provided, but there are no specifics of the data used (for example, not specifying the exact time series). Overcoming this uncertainty in future studies would require attempting to download the exact dataset used in each case, which would be significant undertaking, not least as it would require some subject matter expertise across a diverse range of geophysical subjects. It is worth noting that data repositories do provide application programming interfaces (API) for the datasets which enables programmatic access to items (Figshare). In this work, both a score of 1 or 2 could enable scripted access to data, however for data that score 2 the lack of DOI ultimately means that there is no persistent record. Finally, while the institutional subscriptions that were available to Algarabel and Ireland who undertook the principal data collection provided access to a high proportion of the individual articles reviewed, there were still 27 of 200 articles for which the full text was not accessible.

5 Conclusions

Reproducibility and repeatability are important themes for the geophysics community as evidenced by the increasing number of publications identified in this review. Through examining the current policies of multiple journals which publish geophysical articles, it is identified that all too often the wording used is ambiguous and open to interpretation. If journals want to publish truly reproducible works, it will require not just a shift to using concise wording, but also for journals to enforce stricter policies. Despite this, the empirical evidence is that journals are making a concerted efforts to provide guidance on the provision of data and software. For published articles there are stark differences in the availability and accessibility of both data and software. However, there is still a long way to go for geophysical research (as a whole) to be reproducible, as shown by the findings which indicate that less than 30% of articles over the past 5 years provide enough information on the source of data, and less than 50% of articles identify the software used, both of which are required to reproduce results.

Acknowledgements

The work was enabled by funding from the IAPETUS2 DTP for Guillermo Algarebel to undertake a Research Experience Placement at Newcastle University in the summer of 2022. The authors are grateful for reviews by Daniel Nüst and Wynn Tranfield. Thanks also to Chris Jackson and Matt Hall which stimulated some of the themes discussed in this review.

6 Data and software availability

The data used in this study are available in a repository (Ireland, 2022). There are 5 data tables included, and the description of each is provided below.

- DataTable1_JournalListSciMargo – List of 100 geophysics journals used as starting point for review
- DataTable2_ExistingLiteratureReliability – Number of journal articles published, by year, with the word 'reliability' in the title.
- DataTable3_ExistingLiteratureReproducibility – Number of journal articles published, by year, with the word 'reproducibility' in the title.
- DataTable4_JournalRequirements – Summary of journal requirements categorized.
- DataTable5_PublishedArticles_Anon – Summary of availability of data and software for individual publications. We have removed any identifiable details relating to the individual articles sampled in this study.

The study used the free Publish or Perish software (Harzing, 2007).

All plots were created in Python and the scripts are available (Ireland, 2022). Users will need to download the data dables and add file path locations to the scripts to replicate the plots as they appear in the paper.

Competing interests

The authors have no competing interests.

References

AGU. Data and Software for Authors. https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors#availability. Retrieved 23 October 2022.

American Journal of Political Science. A.J.P.S. Verification Policy, 2019. https://ajps.org/ajps-verification-policy/.

Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R., Krystalli, A., Morley, A., O'Reilly, M., and Whitaker, K. *The Turing Way: A handbook for reproducible data science*. Zenodo, 2019. doi: 10.5281/zenodo.3233853.

Behnke, J., Mitchell, A., and Ramapriyan, H. NASA's Earth Observing Data and Information System – Near-Term Challenges. 18:1, 2019. doi: 10.5334/dsj-2019-040.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, 81(3):530–533, 2010. doi: 10.1785/gssrl.81.3.530.

Boeker, M., Vach, W., and Motschall, E. Google Scholar as replacement for systematic literature searches: Good relative recall and precision are not enough. *BMC Medical Research Methodology*, 13(1):131, 2013. doi: 10.1186/1471-2288-13-131.

Borgman, C. L. Research Data: Who Will Share What, with Whom, When, and Why? *SSRN Electronic Journal*, 2010. doi: 10.2139/ssrn.1714427.

British Geophysical Association. What is geophysics?, 2014. https://geophysics.org.uk/what-is-geophysics/.

Caelleigh, A. S. Role of the journal editor in sustaining integrity in research. *Academic Medicine*, 68(9):23– 29, 1993. doi: 10.1097/00001888-199309000-00030.

Carr, T. R., Buchanan, R. C., Adkins-Heljeson, D., Mettille, T. D., and Sorensen, J. The future of scientific communication in the earth sciences: The impact of the internet. *Computers & Geosciences*, 23(5):503–512, 1997. doi: 10.1016/S0098-3004(97)00032-0.

Childe, S. J. What is the role of a research journal? *Production Planning & Control*, 17(5):439–439, 2006. doi: 10.1080/09537280600888862.

Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., and Rose, A. K. A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PLOS ONE*, 14(12): 0225883, 2019. doi: 10.1371/journal.pone.0225883.

de Groot, P. and Bril, B. The open source model in geosciences and OpendTect in particular. In *SEG Technical Program Expanded Abstracts 2005*, page 802–805. Society of Exploration Geophysicists, 2005. doi: 10.1190/1.2148280.

Dembe, A. E., Partridge, J. S., and Geist, L. C. Statistical software applications used in health services research: Analysis of published studies in the U.S. *BMC Health Services Research*, 11(1): 252, 2011. doi: 10.1186/1472-6963-11-252.

European Commission. G20 Leaders' Communique Hangzhou Summit, 2016. https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967.

Evangelou, E., Trikalinos, T. A., and Ioannidis, J. P. Unavailability of online supplementary scientific information from articles published in major journals. *The FASEB Journal*, 19(14):1943–1944, 2005. doi: 10.1096/fj.05-4784lsf.

Figshare. Figshare API User Documentation. doi: 10.6084/m9.figshare.4880372.v2.

Geophysics. GEOPHYSICS instructions to authors. https://library.seg.org/page/gpysa7/ifa/instructions. Retrieved 23 October 2022.

Glynn, E., Fitzgerald, B., and Exton, C. Commercial adoption of open source software: an empirical study. In *2005 International Symposium on Empirical Software Engineering, 2005*, pages 10 pp.–, 2005. doi: 10.1109/ISESE.2005.1541831.

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M., Smout, C., and Gaynor, K. Why don't we share data and code? *MetaArXiv*, 2022. doi: 10.31222/osf.io/-gaj43.

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. What does research reproducibility mean? *Science Translational Medicine*, 8(341): 341 12–341 12, 2016. doi: 10.1126/scitranslmed.aaf5027.

Hager, B. H. and Clayton, R. W. Constraints on the structure of mantle convection using seismic observations, flow models, and the geoid, 1989. https://resolver.caltech.edu/CaltechAUTHORS:20121002-141328164.

Hamman, J. xarray: N-D labeled Arrays and Datasets in Python, 2017. doi: 10.5334/jors.148.

Harvey, M. J., Mason, N. J., and Rzepa, H. S. Digital Data Repositories in Chemistry and Their Integration with Journals and Electronic Notebooks, 2014. doi: 10.1021/ci500302p.

Harzing, A. W. Publish or Perish. https://harzing.com/resources/publish-or-perish.

Harzing, A. W. *The publish or perish book*. Tarma Software Research Pty Limited Melbourne, Australia, 2010.

Hauge, Ø., Ayala, C., and Conradi, R. Adoption of open source software in software-intensive organizations – A systematic literature review. *Information and Software Technology*, 52(11): 1133–1154, 2010. doi: 10.1016/j.infsof.2010.05.008.

Hodson, S., Jones, S., Collins, S., Genova, F., Harrower, N., Laaksonen, L., Mietchen, D., Petrauskaité, R., and Wittenburg, P. Turning FAIR data into reality. Technical report, 2018. doi: 10.2777/1524. Interim report from the European Commission Expert Group on FAIR data.

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., and Wagenmakers, E. J. Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1):70–85, 2018. doi: 10.1177/2515245917751886.

Ireland, M. Reproducibility in Geophysics, 2022. doi: 10.25405/data.ncl.21564381.v1.

Jun, H. and Cho, Y. Repeatability enhancement of time-lapse seismic data via a convolutional autoencoder. *Geophysical Journal International*, 228(2):1150–1170, 2022. doi: 10.1093/gji/g-gab397.

Konkol, M., Kray, C., and Pfeiffer, M. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science*, 33(2):408–429, 2019. doi: 10.1080/13658816.2018.1508687.

Lepak, D. Editor's Comments: What is Good Reviewing? *Academy of Management Review*, 34(3):375–381, 2009. doi: 10.5465/amr.2009.40631320.

McCullough, B. D. and Heiser, D. A. On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics & Data Analysis*, 52(10):4570–4578, 2008. doi: 10.1016/j.csda.2008.03.004.

McCullough, B. D. and Wilson, B. On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics & Data Analysis*, 40(4):713–721, 2002. doi:

10.1016/S0167-9473(02)00095-6.

McCullough, B. D. and Wilson, B. On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics & Data Analysis*, 49(4):1244–1252, 2005. doi: 10.1016/j.csda.2004.06.016.

Mesirov, J. P. Accessible Reproducible Research. *Science*, 327 (5964):415–416, 2010. doi: 10.1126/science.1179653.

Muenchow, J., Schäfer, S., and Krüger, E. Reviewing qualitative GIS research—Toward a wider usage of open-source GIS and reproducible research practices. *Geography Compass*, 13(6):12441, 2019. doi: 10.1111/gec3.12441.

Mélard, G. On the accuracy of statistical procedures in Microsoft Excel 2010. *Computational Statistics*, 29(5):1095–1128, 2014. doi: 10.1007/s00180-014-0482-5.

National Academies of Sciences. Statistical challenges in assessing and fostering the reproducibility of scientific results: Summary of a workshop, 2016. doi: 10.17226/21915.

National Academies of Sciences. Understanding Reproducibility and Replicability. In *Reproducibility and Replicability in Science*. 2019. https://www.ncbi.nlm.nih.gov/books/NBK547546/. National Academies Press (US).

Nature. Journals unite for reproducibility. *Nature*, 515(7525), 2014. doi: 10.1038/515007a.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., and Christensen, G. Promoting an open research culture. *Science*, 348 (6242):1422–1425, 2015. doi: 10.1126/science.aab2374.

Nüst, D. and Pebesma, E. Practical Reproducibility in Geography and Geosciences. *Annals of the American Association of Geographers*, 111(5):1300–1310, 2021. doi: 10.1080/24694452.2020.1806028.

Oguntimilehin, A. and Ademola, E. O. A Review of Big Data Management, Benefits and Challenges. *A Review of Big Data Management, Benefits and Challenges*, 5(6):6, 2014.

Oren, C. and Nowack, R. L. An overview of reproducible 3D seismic data processing and imaging using Madagascar. *Geophysics*, 83 (2):9– 20, 2018. doi: 10.1190/geo2016-0603.1.

Pendlebury, D. A. The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1):1–11, 2009. doi: 10.1007/s00005-009-0008-y.

Piwowar, H. A., Day, R. S., and Fridsma, D. B. Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLOS ONE*, 2(3):308, 2007. doi: 10.1371/journal.pone.0000308.

Poldrack, R. A. and Gorgolewski, K. J. Making big data open: Data sharing in neuroimaging. *Nature Neuroscience*, 17(11):11, 2014. doi: 10.1038/nn.3818.

Pop, M. and Salzberg, S. L. Use and mis-use of supplementary material in science publications. *BMC Bioinformatics*, 16(1):237, 2015. doi: 10.1186/s12859-015-0668-z.

Rallison, S. What are Journals for? *Annals of The Royal College of Surgeons of England*, 97(2):89–91, 2015. doi: 10.1308/003588414X14055925061397.

Reese, R. J. Recent Applications of Digital Computers to Geophysical Problems. *AAPG Bulletin*, 49(7):1089–1089, 1965. doi: 10.1306/A66336EE-16C0-11D7-8645000102C1865D.

Robinson, E. A. and Treitel, S. *Geophysical signal analysis*. Society of Exploration Geophysicists, 2000.

SCImago. SCImago Journal & Country Rank. https://www.scimagojr.com/aboutus.php. Retrieved 25 August 2022.

Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R., Duerr, R., Haak, L. L., Haendel, M., Herman, I., Hodson, S., Hourclé, J., Kratz, J. E., Lin, J., Nielsen, L. H., Nurnberger, A., Proell, S., Rauber, A., Sacchi, S., Smith, A., and Clark, T. Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*, 1:1, 2015. doi: 10.7717/peerj-cs.1.

Steventon, M. J., Jackson, C. A., Hall, M., Ireland, M. T., Munafo, M., and Roberts, K. J. Reproducibility in subsurface geoscience. *Earth Science, Systems and Society*, 12, 2022. doi: 10.3389/esss.2022.10051.

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., and Sepp, T. Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), 2021. doi: 10.1038/s41597-021-00981-0.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., and Frame, M. Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE*, 6(6):21101, 2011. doi: 10.1371/journal.pone.0021101.

van Rooij, S. W. Higher education sub-cultures and open source adoption. *Computers & Education*, 57(1):1171–1183, 2011. doi: 10.1016/j.compedu.2011.01.006.

Waage, M., Bünz, S., Landrø, M., Plaza-Faverola, A., and Waghorn, K. A. Repeatability of high-resolution 3D seismic data. *Geophysics*, 84(1):75– 94, 2018. doi: 10.1190/geo2018-0099.1.

Walker, R., Gill, S. P., Greenfield, C., McCaffrey, K., and Stephens, T. L. No demonstrated link between sea-level and eruption history at Santorini., 2021. https://eartharxiv.org/repository/view/2638/.

Wildman, G. and Lewis, E. Value of open data: A geoscience perspective. *Geoscience Data Journal*, 2022. doi: 10.1002/gdj3.138.