

Article

Analyzing Data Sharing Policies in Library and Information Science: Journal Metrics, Open Access Status, and Publisher Volume

Eungi Kim , Kristine Joy Tabogoc and Jang Won Chae

Department of Library and Information Science, Keimyung University, 1095 Dalgubeoldae-ro, Dalseo-Gu, Daegu 42601, Republic of Korea; kristinaligayatabogoc@gmail.com (K.J.T.); r2rzzang1@naver.com (J.W.C.)

* Correspondence: egkim@kmu.ac.kr

Abstract: The objective of this study is to explore the prevalence and characteristics of data sharing policies in library and information science (LIS) journals, focusing on their relationship with journal metrics, publisher volume (number of journals managed by the publisher), and open access (OA) status. The study examined the link between data sharing policies and journal metrics in LIS journals indexed in Scopus. Using secondary data from the SJR portal and qualitative data from author guidelines, the study revealed several key findings: First, a positive relationship existed between publisher volume and the presence of data sharing policies, with high-volume publishers consistently implementing such policies. Second, 50.2% of LIS journals lacked data sharing guidelines, indicating a significant gap. Third, journals that encouraged or required data sharing under certain conditions tended to perform better in metrics such as quartiles, h-index, and citation impact. Fourth, data sharing policies were more common in higher-ranked journals and were linked to better journal metrics. Fifth, higher-ranked journals were more likely to include details on data statements, DOIs, and repositories than lower-ranked journals. Lastly, non-OA journals were more likely to encourage practices such as including data repositories and supplementary files in submissions, compared to OA journals, revealing unexpected disparities. In conclusion, data sharing policies in LIS journals showed associations with journal ranking and publisher volume, with non-OA journals showing higher presence of certain data sharing practices.



Citation: Kim, E.; Tabogoc, K.J.; Chae, J.W. Analyzing Data Sharing Policies in Library and Information Science: Journal Metrics, Open Access Status, and Publisher Volume. *Publications* **2024**, *12*, 39. <https://doi.org/10.3390/publications12040039>

Academic Editor: Andrew Kirby

Received: 24 August 2024

Revised: 27 October 2024

Accepted: 1 November 2024

Published: 5 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: data sharing policy; library and information science journals; journal metrics; data availability; journal policy

1. Introduction

There has been a general movement toward sharing data to ensure the transparency of research results in recent decades. This shift is driven by the core principles of the open access (OA) movement, which emphasizes the release of research data to improve transparency, reproducibility, and accountability [1]. In alignment with these principles, some scholarly journals have introduced data sharing policies, reflecting a broader commitment to preserving research integrity [2–4]. By adhering to the principles of OA, scholarly journals contribute to a culture that promotes openness and transparency, leading to the formulation of guidelines and policies that encourage researchers to make their data publicly available [5,6].

Journal guidelines or policies, typically accessible on a journal's homepage, provide various details about manuscript preparation, including instructions on how research data should be accessed and shared with other researchers. Leading international OA journals have increasingly encouraged authors to share their data, underscoring the growing importance of transparent data policies in scholarly publishing. For instance, PLOS One has implemented a data sharing policy that requires authors to include a data availability statement to foster data sharing and research transparency [7], suggesting that OA journals

may be more inclined to adopt data sharing strategies than traditional journals. Moreover, entities like the Committee on Publication Ethics (COPE) have been pivotal in setting ethical standards in journal publishing, guiding stakeholders to engage in transparent and responsible data sharing practices [8].

The principles of OA are particularly significant in LIS [9]. As a dynamic and evolving discipline, LIS emphasizes the importance of understanding issues related to data. This focus ensures that research aligns with ethical transparency standards, fostering credibility and collaboration within the scientific community [10]. Empirical research on the presence of data sharing policies in LIS journals illustrates how the field cultivates a culture of transparency and ethics among researchers. Examining the current state of data sharing policy presence in LIS journals not only strengthens research integrity but also establishes a foundation for continuous advancements in scholarly communication within the field. Moreover, variations in the presence of data sharing policies across publishers of different sizes can reveal recent trends in LIS journal publishing.

Despite the growing adoption of journal data sharing policies and the guidance provided by organizations such as COPE, there is still a notable lack of empirical research examining these policies and their relationship with journal metrics in LIS. This study is important because it provides empirical evidence on the current state of data sharing practices, exploring how they relate to journal metrics, comparing differences between OA and non-OA journals, and investigating how publisher volume (the number of journals managed by the publisher) influences the presence of these policies. The findings from this research can inform publishing strategies, show potential inequalities in the field, and offer valuable information for journal editors in developing data sharing policies. Analyzing data sharing policies in relation to OA status is particularly crucial, as both concepts are grounded in the principles of openness and transparency in scholarly communication. This comparison will help assess the alignment between OA principles and data sharing practices, potentially revealing patterns in how different publishing models approach research transparency.

The objective of this paper is to examine the prevalence and characteristics of data sharing policies in LIS journals, focusing on their association with journal metrics, publisher volume, and OA status. This study also investigates various elements essential to an effective data sharing policy, including guidance provided to authors, the development of detailed data availability statements, and the encouragement of assigning Digital Object Identifiers (DOIs) to datasets and code.

This study addresses the following key research questions:

- How do journals specify and implement data sharing, particularly in their author guidelines and policies?
- How do data sharing policies differ between OA and non-OA LIS journals?
- How do the presence and specific elements of data sharing policies relate to publisher volume and journal metrics, such as h-index and quartile ranking, in LIS?

2. Literature Review

Existing research on data sharing policies paints a complex picture characterized by varying levels of acceptance, challenges, and inconsistencies across disciplines, suggesting concerns about transparency, standardization, and impact on research reproducibility. In particular, previous studies identified challenges in implementing data sharing, including issues related to data privacy, lack of standardized formats, concerns about data misuse, and researchers' general reluctance to disclose sensitive information [11,12]. These studies also addressed data sharing policies related to transparency and managing the large volume of shared health data, as well as the reproducibility of research [13–16]. Vasilevsky et al. [17] found that data sharing policies in biomedical journals vary and often lack specific guidance on maximizing data availability and reusability. These findings demonstrate that implementing data sharing policies among journals is not consistent. Most of these issues predominantly concern biomedical fields, which may differ from interdisciplinary

fields such as LIS. Houtkoop et al. [18] revealed a reluctance to publicly share data in psychology-related research, and Wiley [19] exposed variability in data sharing policies in medical research journals, raising questions about data reusability. Similarly, Jeong [20] found that journals from the Republic of Korea and France were more likely to adopt data sharing policies than those from Brazil, though actual implementation through published data availability statements remained uncommon.

Previous studies have also found disciplinary differences in the approach to data sharing policies. Rousi and Laakso [21] identified significant variations in data sharing policies among highly cited journals in neuroscience, physics, and operations research, noting a notable influence from the inclusion of life science data types. They offered open data and a coding framework for ongoing research and policy monitoring in the realm of open science. Wang et al. [22] examined the research data policies of Chinese scholarly journals, revealing that these policies were generally weak, with variations based on factors such as journal language, publisher type, discipline, access model, and journal metrics. In a related context, Kim et al. [23] revealed that 44.0% of 700 journals in life, health, and physical science lacked a data sharing policy, while 38.1% had a strong policy mandating or expecting data sharing. In particular, English-language journals, journals co-published by Western and Chinese publishers, life science journals, and journals with higher impact factors can be observed to have stronger data policies. Tal-Socher [24] examined data sharing policies across disciplines and found that they vary in prevalence, with the biomedical sciences leading the way and the humanities lagging behind.

Several studies have explored the connection between data policy and journal impact, revealing emerging patterns in scholarly communication. Kim et al. [23] found that higher impact factor quartiles were positively associated with stronger data sharing policies. Kim and Bai [25] reported that approximately 29.5% of the Asian journals they studied had data sharing policies, with significant correlations between these policies, impact factor, and commercial publisher type. The authors noted that while many journals encouraged data sharing, more effective strategies are needed to ensure compliance and explore varying policy strengths in Asian journals.

In LIS, prior studies have addressed data sharing policies. Alexandre-Benavent et al. [26] found a positive relationship between being a top-ranked journal in the Journal Citation Reports (JCR) and having an open policy, with a significant percentage of LIS journals supporting data reuse, storage in repositories, and public availability of research data. In addition, Jackson [3] analyzed open data policies among LIS journals, and Thøgersen and Borlund [4] conducted a meta-evaluation of studies on researcher attitudes toward data sharing in public repositories.

Although the above-mentioned previous studies offer valuable perspectives, there remains a need for more comprehensive and up-to-date research in LIS journals, particularly in relation to journal metrics, OA status, and publisher volume. This study aims to address this gap by providing a broader and more current analysis of data sharing practices in Scopus-indexed LIS journals. Specifically, this research explores the differences in data sharing policies between OA and non-OA journals and examines how these practices relate to publisher volume and various journal metrics.

3. Methodology

The study examined the extent and association between the presence of data sharing policies and journal metrics in LIS journals indexed in Scopus for the year 2023. Initially, 262 journals categorized under LIS in the SCImago portal were considered. Two types of data were utilized: (a) secondary data on journal indicators and (b) qualitative data manually collected from the journals' websites. The qualitative data were gathered by examining online author statements and editorial policies on the websites of LIS journals. These two data types were analyzed together to evaluate the current landscape of the presence of data sharing policies and their association with journal metrics. Out of the 262 LIS journals initially considered, 31 journals that lacked clear and detailed author

submission guidelines were excluded. Consequently, following this exclusion, 231 journals were retained for this study. We used the R programming language to merge journal metrics with qualitative data and analyze the research data, which were collected in July 2024.

To qualitatively assess the data sharing policies, we manually reviewed the journal policies and author information available on each journal's homepage. Links to the journal homepages were obtained from the SCImago portal, and in cases where the link was unclear, an internet search engine was used to locate the correct site. The data sharing policies were evaluated based on predefined criteria, and any discrepancies in interpretation were resolved to maintain consistency. The qualitative data, including the prevalence and types of data availability statements in the author guidelines sections, were systematically analyzed to ensure a comprehensive review.

3.1. Data Collection of Journal Metrics and Attributes

This study utilized journal metrics data sourced from the SCImago portal. The key metrics analyzed include the SCImago Journal Rank (SJR) [27], h-index, the total number of documents produced by the journal in 2023 (TD2023), and the average number of citations per document over a two-year period (CITES2YR). Further details and definitions for these indicators are available on the SCImago portal [28]. Additionally, the study examined the presence of data sharing policies in relation to the journal's OA status (OA versus non-OA), as identified through the Directory of Open Access Journals (DOAJ) and the Directory of Open Access Scholarly Resources (ROAD).

3.2. Qualitative Data Collection

For the 231 LIS journals with clear manuscript submission guidelines on their websites, qualitative data were collected based on two key aspects of data sharing policies (Table 1). We categorized journals according to their approach to data sharing: "No Mention" for journals that do not reference data sharing, "Encourage/Expect" for those that promote or anticipate data sharing without requiring it, and "Conditional" for journals that require authors to provide data if requested as a condition of publication. Although these categories align closely with prior research (e.g., Kim et al. [23]; Jeong. [20]), we made slight adaptations to capture nuances specific to LIS journals. This tailored framework reflects both the general structure seen in previous studies and the particular practices we observed, ensuring a more accurate representation of data sharing policies within this field.

In addition to policy categories, we examined key policy elements aligned with best practices in scholarly publishing, including data availability statements, repository specifications, DOI assignment for datasets or code, and the inclusion of supplementary files. These elements reflect essential components that promote transparency and reproducibility in research. Each element was coded in binary terms ("Yes" or "No") to indicate whether the journal explicitly addressed it in its guidelines. This binary approach ensures clarity and consistency in the analysis. Throughout the data collection process, we carefully reviewed and cross-checked the encoding to minimize errors and ensure consistency. This structured process ensures the accuracy of our coding and supports the replicability of the study's findings, even with the nuanced differences in our adapted framework.

Table 2 presents example statements of data sharing policy categories and key elements of data sharing policies as outlined in journal guidelines. The table includes statements that encourage, expect, or conditionally mandate data sharing, alongside specific components such as data availability statements, repository use, DOI assignment, and the submission of supplementary files. These examples demonstrate how journals establish guidelines to promote transparency and accountability in research practices.

Table 1. Description and options of elements of data sharing policy.

Aspects	Variable	Description	Options
Categories of Data Sharing Policy	Data Sharing Policy Categories (policy_category)	No Mention: The journal’s information for authors or journal policy statements do not make any references to expectations or requirements for sharing data. Encourage/Expect: The journal explicitly suggests or anticipates authors share their research data as part of the publication process, though it may not be compulsory. Conditional: Authors are required to provide data upon request as a condition for publication.	No Mention; Encourage/Expect; Conditional
Elements of Data Sharing Policy	Data Availability Statement (data_statement)	Requests authors to specify the availability and access details of the data used in their research, formally outlining how their data can be accessed if they choose to share it.	Yes/No
	Repository Specification (repo_spec)	Mentions that authors can use repositories for sharing their research data.	Yes/No
	DOI for Data and Code (doi_data_code)	Encourages assigning a Digital Object Identifier (DOI) to datasets and code.	Yes/No
	Supplementary Files (supp_files)	Explicitly specifies requirements for providing additional data files related to the research for data sharing purposes, including preparation and submission details.	Yes/No

Table 2. Example statements of categories and elements of data sharing policy.

Item	Example Statement
Data Sharing Policy Categories (policy_category)	Encourage: “Authors are encouraged to share their data in order to enhance transparency in research”.
	Expect: “Authors may be requested to provide information on how their research data can be accessed and cited in their manuscripts”.
	Conditional: “Authors may be required to provide data upon request as a condition for publication”.
Data Availability Statement (data_statement)	“Authors are encouraged to include a statement on how their data can be accessed, specifying the availability and access details of the data used in their research”.
Repository Specification (repo_spec)	“Authors can make their data and software code available to readers by depositing it in a public repository such as [repository name], where ethically feasible”.
DOI for Data and Code (doi_data_code)	“Authors are encouraged to assign a DOI to their datasets, ensuring that this journal can maintain long-term accessibility and proper citation standards”.
Supplementary Files (supp_files)	“[The name of the journal] allows authors to submit supplementary files to support their manuscript, enhancing the transparency and reproducibility of their research”.

3.3. Data Cleaning and Processing

Building on the qualitative data collection, further data processing was conducted to facilitate in-depth analysis. Journals were categorized based on their data sharing policies—“No Mention”, “Encourage/Expect”, and “Conditional”—and key elements of data sharing policies—data_statement, doi_data_code, repo_spec, and supp_files—were combined with journal attributes and metrics. The dataset was then refined by adding a variable indicating the journal type (OA vs. non-OA), which was used in subsequent analyses.

To explore the relationship between publisher volume and the presence of data sharing policies, publishers were grouped by the number of journals they manage. This “publisher volume” provided a straightforward and quantifiable metric for analysis. A moderate level of data cleaning and normalization was performed to ensure accuracy and comparability. These steps included standardizing publisher names and removing redundant suffixes (e.g., “AG”, “Inc.”, and “Ltd.”). The final dataset incorporated the presence of data sharing policies across various publisher volumes, forming a reliable basis for statistical analyses. Given the complexities involved in measuring publisher size, publisher volume—reflecting the number of journals managed by each publisher—was selected as a relevant and consistent metric for this study. For example, a publisher managing 25 journals would have a publisher volume of 25, while one managing a single journal would have a publisher volume of 1.

3.4. Statistical Analysis

In this study, several statistical analyses were selected to accurately assess the relationships between data sharing policies and various journal metrics in LIS journals. Mann–Whitney U tests were chosen to compare journal metrics across different data sharing policy categories because this non-parametric test is suitable for comparing independent samples without assuming normal distribution. Chi-square tests were applied to evaluate the association between categorical variables, such as journal quartiles and specific elements of data sharing policies, as they effectively analyze relationships between categorical data. Furthermore, ordinal logistic regression was employed to explore the relationship between publisher volume and journal quartile rankings, given its robustness in modeling ordered outcomes like quartile rankings. These statistical approaches were chosen to ensure a comprehensive and rigorous analysis, revealing the landscape of data sharing policies across LIS journals and their association with journal metrics and attributes.

4. Results

4.1. Distribution of Data Sharing Policy Categories

Data sharing policies were analyzed with respect to different policy categories mentioned previously (Table 3). As shown, the majority of journals (50.2%) fall into the “No Mention” category. This indicates that a significant portion of journals do not explicitly address or require specific data sharing in their author guidelines, suggesting a lack of emphasis on data sharing policies within these journals. On the other hand, a substantial amount of journals (48.5%) are categorized under “Encourage/Expect”, indicating that a sizable portion of the academic community promotes and supports data sharing practices or at least expects researchers to consider it. Encouragement or expectation regarding data sharing by LIS journals reflects a recognition of the importance of sharing research data for transparency, reproducibility, and collaborative research efforts. The smallest subset of journals falls into the “Conditional” category (1.3%), representing journals that require researchers to share their data as a condition of publication, indicating a formalized approach to data sharing practices.

Table 3. Journal distribution across policy categories.

No Mention	Encourage/Expect	Conditional	Total
116 (50.2%)	112 (48.5%)	3 (1.3%)	231 (100%)

4.2. Distribution of Data Sharing Policy Elements

We also analyzed data sharing policies by focusing on key elements such as the inclusion of a formal data availability statement (`data_statement`), the use of DOIs for datasets (`doi_data_code`), the specification of repositories for data storage (`repo_spec`), and the provision of supplementary files (`supp_files`). Figure 1 displays the frequency distribution of the presence of these elements among LIS journals. Regarding the `data_statement` variable,

72 journals (31.17%) included a data statement, while 159 journals (68.83%) did not. For doi_data_code, 76 journals (32.9%) explicitly mentioned a DOI for data sharing, whereas 155 journals (67.1%) did not. Similar patterns are observed for the other two variables, repo_spec and supp_files. Specifically, 105 journals (45.45%) mentioned that authors could use repositories for sharing their research data (repo_spec), and 111 journals (48.05%) explicitly specified requirements for providing supplementary files related to the research (supp_files), including preparation and submission details. These findings suggest that while elements like repo_spec and supp_files are relatively common in LIS journals, there is less emphasis on the data_statement, which requests authors to specify how and where supporting data can be accessed if they choose to share them. This indicates that many LIS journals may prioritize other aspects of data sharing over formalizing the accessibility of supporting data, possibly due to varying editorial practices or resource limitations.

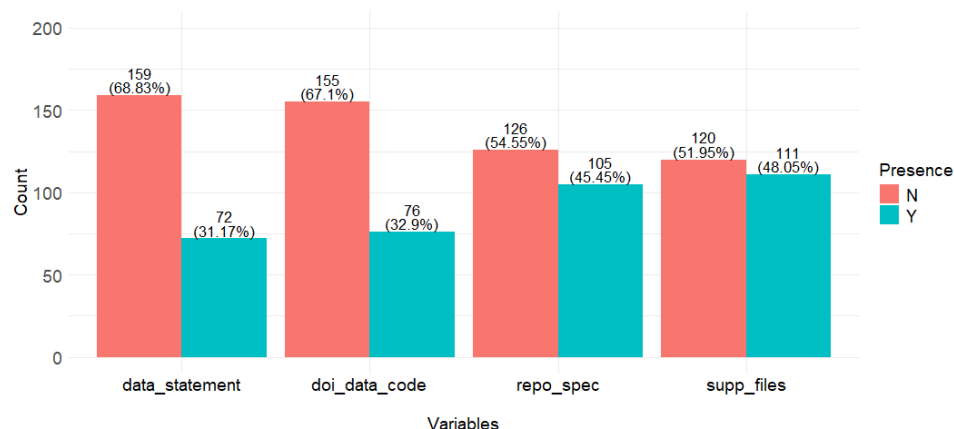


Figure 1. Distribution of data sharing policy elements in LIS journals.

4.3. Presence of Data Sharing Policy and Journal Metrics

Table 4 presents the results of Mann–Whitney U tests comparing journal metrics between those that do not mention data sharing policies (“No Mention”) and those that encourage or expect such policies (“Encourage/Expect” or “Conditional”). The analysis reveals significant differences across all metrics, with *p*-values well below 0.05. For SJR, journals that encourage or expect data sharing have a higher mean value of 0.721 compared to 0.326 for those that do not, with a *W* statistic of 2319.5. The *h*-index is also higher in the “Encourage/Expect” group, with a mean of 44.504 versus 20.491, supported by a *W* statistic of 2813.5. The TD2023 metric shows a mean of 84.800 for journals that promote data sharing, compared to 30.948 for those that do not, with a *W* statistic of 3664.0. Lastly, the CITES2YR metric, reflecting citations over two years, is higher in the “Encourage/Expect” group with a mean of 2.904 compared to 0.907, as indicated by a *W* statistic of 2267.0. These results indicate that journals with policies encouraging or expecting data sharing tend to exhibit higher performance metrics.

Table 4. Mann–Whitney U test results for journal metrics and presence of data sharing policies.

Variable	W_Statistic	<i>p</i> -Value	Mean Value (No Mention)	Mean Value (Encourage/Expect or Conditional)
SJR	2319.5	1.78×10^{-17} ***	0.326	0.721
<i>h</i> -index	2813.5	3.07×10^{-14} ***	20.491	44.504
TD2023	3664.0	3.22×10^{-9} ***	30.948	84.800
CITES2YR	2267.0	4.35×10^{-18} ***	0.907	2.904

Note: *** *p* < 0.0001. The significance level for hypothesis testing is set at 0.05.

To examine the association between data sharing policies and journal quartiles, we analyzed the distribution of specific data policy elements across different journal rankings (Figure 2). For data available statements, there is a clear downward trend, with the proportion of journals including a data statement decreasing from 49.3% in Q1 to 15.0% in Q4. The doi_data_code element shows a similar downward pattern, decreasing from 49.3% in Q1 to 15.0% in Q4, indicating that higher-ranking journals are more likely to include DOIs for research data compared to lower-ranking journals. The repo_spec element shows a decreasing trend across quartiles. Specifically, 41.7% of Q1 journals specify a repository, decreasing to 37.5% in Q2, 22.0% in Q3, and 15.0% in Q4. A similar pattern is observed for supp_files, with 86.6% of Q1 journals including supplementary files, compared to 15.0% in Q4. These trends suggest that higher-ranked journals tend to adopt key data sharing practices more frequently, such as specifying repositories or including supplementary files, while lower-ranked journals are less consistent in their implementation. Although these differences reflect some variability in how journals approach data sharing, they may also point to resource or policy differences between journal tiers.

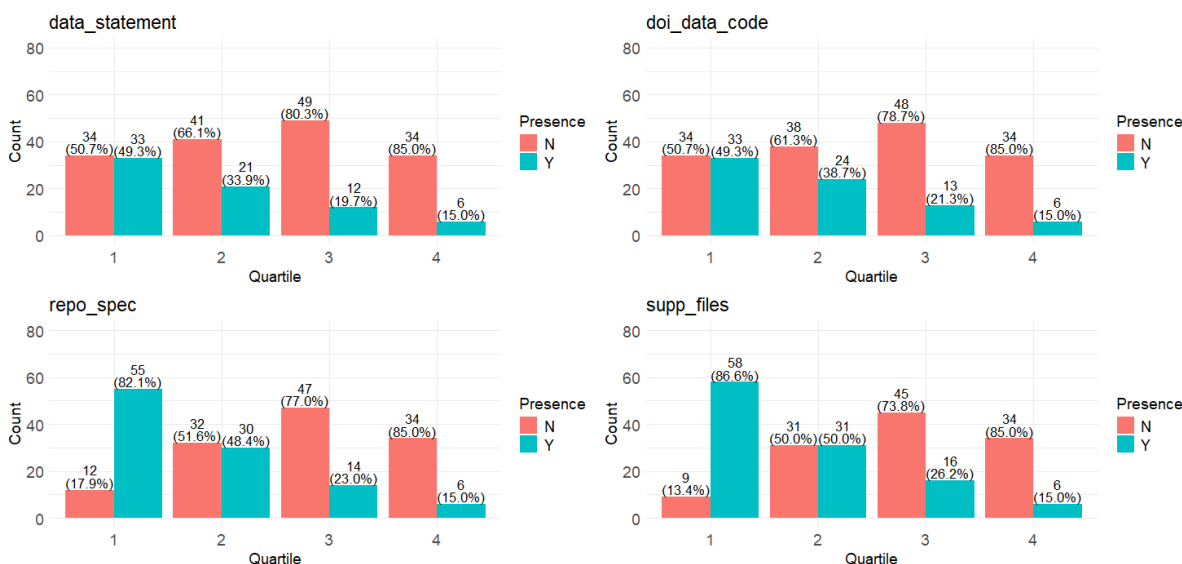


Figure 2. Presence of data sharing policy elements across journal quartiles.

Table 5 presents the results of chi-square tests examining the association between journal quartiles and various data sharing policy elements. The chi-square statistics and corresponding *p*-values indicate statistically significant associations for all variables, with *p*-values well below 0.0003. Specifically, the presence of data_statement, doi_data_code, repo_spec, and supp_files are all significantly associated with journal quartiles, as indicated by the chi-square statistics ranging from 18.5 to 69. These results suggest a strong association between journal quartiles and the presence of data sharing policies, indicating that higher-ranked journals may be more inclined to implement and emphasize these practices.

Table 5. Chi-square test results for journal quartiles and data sharing policies.

Variable	Description	Chi-Square Statistic	df	<i>p</i> -Value
data_statement	Presence of a data availability statement	19	3	$2.72 \times 10^{-4} *$
doi_data_code	DOI provided for datasets	18.5	3	$3.41 \times 10^{-4} *$
repo_spec	Specified repository for data storage	63.9	3	$8.8 \times 10^{-14} *$
supp_files	Availability of supplementary files	69	3	$8.8 \times 10^{-14} *$

Note: * *p* < 0.05 indicates statistical significance.

4.4. Elements of Data Sharing Policy and OA Status of Journal

Figure 3 illustrates the relationship between different elements of data sharing policies and the OA status of journals. For `data_statement`, 34.59% of non-OA journals explicitly describe whether and how the data supporting the findings can be accessed, compared to 23.61% of OA journals. For `doi_data_code`, 37.11% of non-OA journals encourage assigning a DOI to datasets and code, compared to 23.61% of OA journals. Regarding `repo_spec`, 54.09% of non-OA journals mention that authors can use repositories for sharing their research data, while 26.39% of OA journals include this element. For `supp_files`, 57.23% of non-OA journals explicitly specify requirements for providing supplementary data files, compared to 27.78% of OA journals. Overall, non-OA journals tend to adopt these data sharing policy elements at a higher rate compared to OA journals. An interesting trend to note is that the `supp_files` element is the most prevalent among both non-OA (57.23%) and OA journals (27.78%), suggesting that the provision of supplementary files is a significant practice across different journal types, regardless of OA status.

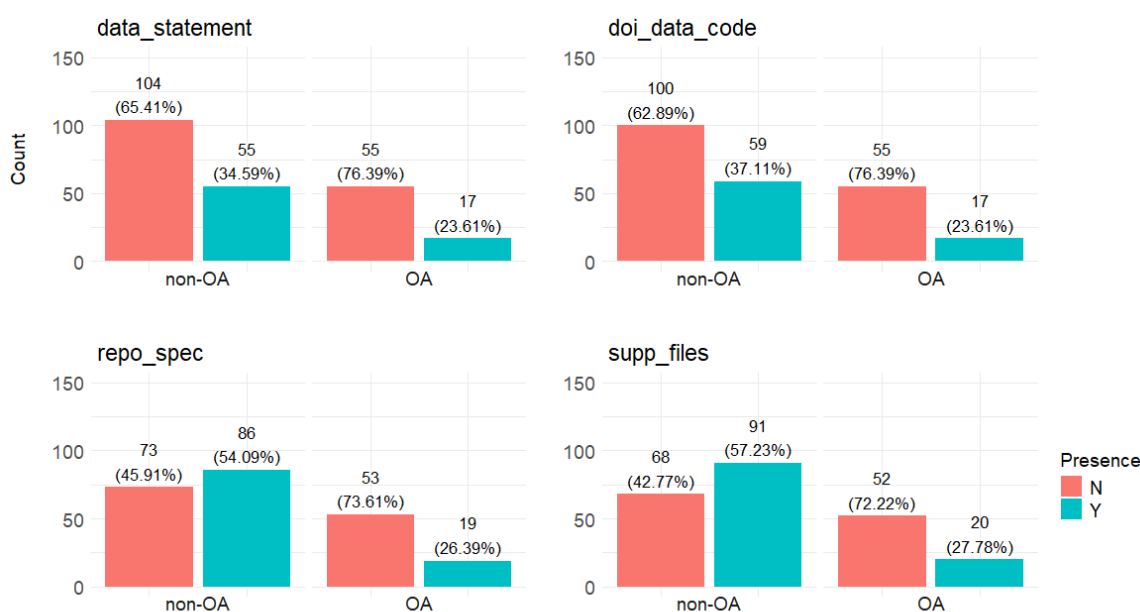


Figure 3. Elements of data sharing policy by OA status of journals (non-OA and OA). Note: OA vs. non-OA journals—`data_statement`: chi-square = 2.297, $p = 0.13$ (no significant difference); `doi_data_code`: chi-square = 3.500, $p = 0.0614$ (no significant difference); `repo_spec`: chi-square = 14.239, $p = 1.61 \times 10^{-4}$ (significant difference); `supp_files`: chi-square = 16.065, $p = 6.12 \times 10^{-5}$ (significant difference).

As shown in the note section in Figure 3, chi-square tests comparing OA and non-OA journals revealed no significant differences for the variables `data_statement` (chi-square = 2.297, $p = 1.30 \times 10^{-1}$) and `doi_data_code` (chi-square = 3.500, $p = 6.14 \times 10^{-2}$). However, significant differences were found for `repo_spec` (chi-square = 14.239, $p = 1.61 \times 10^{-4}$) and `supp_files` (chi-square = 16.065, $p = 6.12 \times 10^{-5}$), indicating that OA and non-OA journals differ significantly in their practices regarding repository specification and the inclusion of supplementary files. This suggests that OA and non-OA journals are similar in terms of including data statements and providing DOIs for data and code. However, non-OA journals may have more established practices for repository use and supplementary material inclusion.

4.5. Presence of Data Sharing Policy and Publisher Volume

4.5.1. Relationship Between Publisher Volume and the Presence of Data Sharing Policies

Figure 4 illustrates the relationship between publisher index, publisher volume, and the presence of data sharing policies in the author guidelines for LIS journals indexed in

Scopus. In the 2023 indexing, we identified a total of 129 publishers, with some managing more than one journal. Of these, 113 publishers manage only a single journal, and the majority (85 out of 113) do not have a data sharing policy, while 28 have such a policy in place. For publishers managing two to five journals, there is a mixed pattern regarding the presence of data sharing policies. However, publishers with larger volumes (those managing more than five journals) consistently implement data sharing policies, with every publisher in this category having such policies. This analysis suggests a positive relationship between publisher volume and the likelihood of adopting a data sharing policy. While the data indicate that higher-volume publishers are more likely to standardize such policies, further investigation would be needed to determine the underlying factors, such as available resources or infrastructure.



Figure 4. Publisher index, volume, and data sharing policy presence. Note: Each ‘P#’ represents a publisher, with the number indicating how many journals they manage (e.g., P25 manages 25 journals). Numbers between indices may be missing when no publisher manages that specific number of journals (e.g., there is no P24 because no publisher in our dataset manages exactly 24 journals).

4.5.2. Association Between Publisher Volume and Journal Quartile Ranking

We conducted an ordinal logistic regression analysis to examine the association between publisher volume (*pub_vol*) as the independent variable and journal quartile ranking (*QUAR*) as the dependent variable. Understanding how publisher volume relates to quartile rankings is important, as higher-ranked journals tend to demonstrate greater transparency in their policies, including data sharing practices. In addition, publishers with larger portfolios may have more resources and infrastructure, which can facilitate the adoption of standardized policies across their journals.

The model results (Table 6) show a significant positive association between publisher volume and quartile ranking, with an estimated coefficient of 0.06249 (SE = 0.01398, $z = 4.471$, $p = 7.79 \times 10^{-5}$). This indicates that as the number of journals managed by a publisher increases, the likelihood of a journal being classified into a higher quartile also increases, suggesting that larger publishers are more likely to manage higher-impact journals.

The model’s threshold coefficients represent the cut-off points between adjacent quartiles. While the thresholds for the 4|3 and 2|1 transitions ($z = -6.376$ and $z = 7.318$, respectively) indicate clear distinctions between these quartiles, the 3|2 threshold ($z = 1.072$) is not statistically significant. This suggests that publisher volume alone does not provide a strong distinction between journals in Q2 and Q3.

Table 6. Ordinal logistic regression results: publisher volume and journal quartile.

Effect	Term	Estimate	Std. Error	z value	p-value
	pub_vol	0.06249	0.01398	4.471	7.79×10^{-5}
	Threshold	Estimate	Std. Error	z value	
Threshold	4 3	−1.2889	0.1940	−6.376	
	3 2	0.1798	0.1677	1.072	
	2 1	1.4272	0.1950	7.318	

Note. The z-score for pub_vol is 4.471 ($p = 7.79 \times 10^{-5}$), indicating a significant relationship between publisher volume and higher journal quartile rankings. The threshold z-scores are −6.376 (4|3), 1.072 (3|2), and 7.318 (2|1), showing significant separation between quartiles 4|3 and 2|1, but less distinction between quartiles 3|2.

The similarity between Q2 and Q3 journals could result from metric overlap—such as similar citation counts, h-indexes, or SJR scores—which causes journals to cluster near the boundary between these quartiles. Additionally, volatility in annual citation metrics could cause journals near the Q2–Q3 boundary to shift between quartiles over time, further blurring the distinction. This overlap suggests that publisher volume alone is insufficient to predict quartile distinctions in the middle tiers, highlighting the complexity of ranking metrics.

Overall, these results suggest that as publisher volume increases, the likelihood of a journal being in either the top (Q1) or bottom (Q4) quartile becomes more pronounced. However, the distinction between journals in the middle quartiles (Q2 and Q3) is less clear, indicating that these journals may share more similar characteristics compared to those at the extremes. These findings highlight the importance of understanding how publisher practices and resources might influence both quartile rankings and the adoption of transparent policies, such as data sharing.

5. Discussion

In this study, for the qualitative data processing, we began by categorizing data sharing policies and encountered difficulties in applying previous frameworks that used hierarchical levels (as seen in [23]), as they often lacked clear distinctions. Our refined categorization—“No Mention”, “Encourage/Expect”, and “Conditional”—provides a more straightforward approach to assessing journal commitment to data transparency. The “No Mention” category reflects the absence of guidance on data sharing, leading to inconsistencies in authors’ practices and potential reductions in research transparency. “Encourage/Expect” indicates a journal’s support for data sharing, which can be explicit or implicit, allowing authors some discretion and resulting in varying compliance levels. The “Conditional” category requires authors to provide data upon request as a condition for publication, adding accountability but also introducing uncertainty, as data sharing is not an absolute requirement. These categories underscore the challenges in standardizing data sharing practices across journals, as each approach strikes a different balance between flexibility and accountability.

Some of the findings of this study closely mirror those previously reported in other disciplines. For example, we identified a notable deficiency in the presence of proactive measures for data sharing policies, with a substantial portion (50.2%) of LIS journals failing to address any type of data sharing policy. This is comparable to findings by Kim et al. [23], who reported that 44% of journals in life, health, and physical sciences lacked a data sharing policy. Furthermore, our investigation revealed that LIS journals with explicit data sharing policies tended to have improved journal metrics and rankings. These outcomes are consistent with prior studies, which affirm that the presence of data sharing policies contributes to increased journal impact [23–25].

This study finds a statistically significant link between the presence of data sharing policies and higher journal metrics, particularly in top-ranked journals (Q1). Journals with these policies tend to have better scores in metrics like SJR, h-index, TD2023, and CITES2YR,

suggesting that encouraging data sharing is associated with higher journal metrics. The analysis, using chi-square tests, shows that leading journals are more likely to promote transparency and research integrity by including data sharing in their guidelines. This trend underscores the importance of data sharing for enhancing credibility and reproducibility in research, especially in LIS journals.

Although OA journals are typically associated with openness and transparency, exemplified by PLOS One's data sharing requirements, our results reveal a different pattern. Our analysis shows that non-OA journals are more likely to adopt specific elements of data sharing policies, particularly regarding repository specifications and supplementary files, where we found statistically significant differences. This finding challenges the assumption that OA journals, grounded in the principles of open access, would naturally lead in promoting data sharing. These results suggest that the relationship between OA status and data sharing practices may be more complex than initially assumed. One possible explanation is that OA journals, on average, may be smaller in terms of publisher volume, which could influence the scope and consistency of their data sharing policies. Further research is needed to confirm and refine these findings, particularly by investigating how OA and non-OA journals differ across varying publisher volumes and how these differences shape data sharing practices.

Our findings further show the importance of publisher volume in shaping data sharing policies. Low-volume publishers often lack explicit policies, whereas high-volume publishers are more likely to explicitly mention data sharing policies for authors. Our analysis showed significant relationships between publisher volume and quartile rankings, though this relationship was less pronounced between middle-tier journals. Separately, our analysis revealed that larger publishers consistently implemented data sharing policies, with all publishers managing more than five journals having such policies in place. Together, these patterns indicate that larger publishers are associated with both higher-ranked journals and more consistent implementation of data sharing policies, which may reflect their available infrastructure. Low-volume publishers may face limitations in allocating resources to the development and implementation of explicit data sharing policies. Furthermore, a general positive association was found between the presence of a data sharing policy and journal metrics, suggesting that journals managed by high-volume publishers may take a more systematic approach to the publishing process by addressing various aspects such as transparency, ethical standards, and data sharing practices.

There are several limitations to this study that are worth acknowledging. Firstly, our analysis focused on how journals address data sharing policies within their author guidelines, rather than examining the actual data sharing practices of authors or their adherence to these guidelines. The extent to which authors make data available might differ from the journal's policy, as pointed out in prior studies [29]. Secondly, there is a temporal gap between the collection of data sharing policy information and the journal data from SJR (i.e., indexed for 2023 but collected in July 2024). Lastly, we acknowledge the evolving nature of data sharing policy in scholarly journals. Our study captures a momentary overview of current practices, but journals can introduce new data sharing policies or alter author guidelines at any point. This dynamic nature of policy changes and modifications could influence the interpretation of our results and their applicability over time.

6. Conclusions and Recommendations

Our study found clear relationships between data sharing policies in LIS journals and their metrics, publisher volume, and rankings. Higher-ranked journals and larger publishers were more likely to adopt transparent data sharing practices, reflecting the connection between institutional resources and policy implementation. While journals with data sharing policies showed improved metrics (higher SJR scores, h-indexes, and citation impact), we emphasize that ethical principles, not just metrics, should guide policy decisions [10,11,30].

Despite the growing importance of data sharing, practical barriers remain. Researchers have raised concerns about privacy, data misuse, and the time required to prepare datasets for public release [31]. Journals must address these issues by offering clear guidance, resources, and incentives for data sharing. Although curating data is time-consuming [32], it brings long-term benefits for transparency and collaboration. Journals should adopt stakeholder-driven data sharing models to ensure editors, reviewers, and authors follow best practices [33]. These models would support consistent and reliable data sharing practices across all fields, including LIS.

The most interesting finding of this study is the positive relationship between publisher volume and the presence of data sharing policies. Our analysis shows that high-volume publishers, with more resources, have consistently implemented comprehensive data sharing policies, while low-volume publishers often lack these policies. This disparity suggests a structural challenge in scholarly publishing where resource differences between high-volume and low-volume publishers may affect not only journal rankings but also their ability to implement standardized data sharing practices. This finding has important implications for understanding how publisher characteristics influence scholarly communication practices.

Future studies could examine differences in data sharing practices across LIS subfields and regional journals [34]. This would reveal patterns and cultural differences in data sharing norms, helping journals improve transparency. Future studies could further explore how publishing models shape data sharing practices, uncovering the connections between business strategies and research openness. Ultimately, data sharing should become a core part of scholarly communication, independent of publishing models.

Author Contributions: Conceptualization, E.K., K.J.T. and J.W.C.; methodology, E.K.; validation, E.K., K.J.T. and J.W.C.; formal analysis, E.K.; investigation, E.K., K.J.T. and J.W.C.; resources, E.K., K.J.T. and J.W.C.; data curation, E.K., K.J.T. and J.W.C.; writing—original draft preparation, E.K.; writing—review and editing, E.K., K.J.T. and J.W.C.; visualization, E.K.; supervision, E.K. project administration, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The journal metrics data are publicly available from the SCImago portal (<https://www.scimagojr.com/>, accessed on 31 July 2024). The qualitative data supporting this study will be made available by the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Elliott, K.C.; Resnik, D.B. Making open science work for science and society. *Environ. Health Perspect.* **2019**, *127*, 075002. [CrossRef] [PubMed]
2. Golub, K.; Hansson, J. (Big) data in library and information science: A brief overview of some important problem areas. *J. Univers. Comput. Sci.* **2017**, *23*, 1098–1108. [CrossRef]
3. Jackson, B. Open data policies among library and information science journals. *Publications* **2021**, *9*, 25. [CrossRef]
4. Thøgersen, J.L.; Borlund, P. Researcher attitudes toward data sharing in public data repositories: A meta-evaluation of studies on researcher data sharing. *J. Doc.* **2022**, *78*, 1–17. [CrossRef]
5. Wilkinson, M.; Amos, H.; Morton, L.; Flaherty, B.; Hearne, S.; Lynch, H.; Lamond, H.; Dewson, N.; Kmiec, M.; Nicolle, J.; et al. *Research Data Management Framework Report*; CONZUL Working Group: Wellington, New Zealand, 2016.
6. Hicks, D.J. Open science, the replication crisis, and environmental public health. *Account. Res.* **2023**, *30*, 34–62. [CrossRef]
7. Jiao, C.; Li, K.; Fang, Z. Data sharing practices across knowledge domains: A dynamic examination of data availability statements in *PLoS ONE* publications. *J. Inf. Sci.* **2022**, *50*, 673–689. [CrossRef]
8. COPE. Promoting Integrity in Research and Its Publication. Available online: <https://publicationethics.org> (accessed on 13 July 2024).
9. Pilerot, O. LIS research on information sharing activities—people, places, or information. *J. Doc.* **2012**, *68*, 559–581. [CrossRef]
10. Bauchner, H.; Golub, R.M.; Fontanarosa, P.B. Data sharing: An ethical and scientific imperative. *JAMA* **2016**, *315*, 1238–1240. [CrossRef]
11. Boté, J.J.; Termens, M. Reusing data: Technical and ethical challenges. *DESIDOC J. Libr. Inf. Technol.* **2019**, *39*, 6. [CrossRef]

12. Kaye, J. The tension between data sharing and the protection of privacy in genomics research. *Annu. Rev. Genom. Hum. Genet.* **2012**, *13*, 415–431. [[CrossRef](#)]
13. Bergeat, D.; Lombard, N.; Gasmi, A.; Le Floch, B.; Naudet, F. Data sharing and reanalyses among randomized clinical trials published in surgical journals before and after adoption of a data availability and reproducibility policy. *JAMA Netw. Open* **2022**, *5*, e2215209. [[CrossRef](#)] [[PubMed](#)]
14. Lombard, N.; Gasmi, A.; Sulpice, L.; Boudjema, K.; Naudet, F.; Bergeat, D. Research transparency promotion by surgical journals publishing randomised controlled trials: A survey. *Trials* **2020**, *21*, 824. [[CrossRef](#)] [[PubMed](#)]
15. Siebert, M.; Gaba, J.F.; Caquelin, L.; Gouraud, H.; Dupuy, A.; Moher, D.; Naudet, F. Data sharing recommendations in biomedical journals and randomised controlled trials: An audit of journals following the ICMJE recommendations. *BMJ Open* **2020**, *10*, e038887. [[CrossRef](#)]
16. Waithira, N.; Mutinda, B.; Cheah, P.Y. Data management and sharing policy: The first step towards promoting data sharing. *BMC Med.* **2019**, *17*, 80. [[CrossRef](#)]
17. Vasilevsky, N.A.; Minnier, J.; Haendel, M.A.; Champieux, R.E. Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ* **2017**, *5*, e3208. [[CrossRef](#)]
18. Houtkoop, B.L.; Chambers, C.; Macleod, M.; Bishop, D.V.; Nichols, T.E.; Wagenmakers, E.J. Data sharing in psychology: A survey on barriers and preconditions. *Adv. Methods Pract. Psychol. Sci.* **2018**, *1*, 70–85. [[CrossRef](#)]
19. Wiley, C. Data sharing: An analysis of medical faculty journals and articles. *Sci. Technol. Libr.* **2021**, *40*, 104–115. [[CrossRef](#)]
20. Jeong, G.H. Status of the Data Sharing Policies of Scholarly Journals Published in Brazil, France, and Korea and Listed in Both the 2018 Scimago Journal and Country Ranking and the Web of Science. *Sci. Ed.* **2020**, *7*, 136–141. [[CrossRef](#)]
21. Rousi, A.M.; Laakso, M. Journal research data sharing policies: A study of highly-cited journals in neuroscience, physics, and operations research. *Scientometrics* **2020**, *124*, 131–152. [[CrossRef](#)]
22. Wang, Y.; Chen, B.; Zhao, L.; Zeng, Y. Research data policies of journals in the Chinese Science Citation Database based on the language, publisher, discipline, access model, and metrics. *Learn. Publ.* **2022**, *35*, 30–45. [[CrossRef](#)]
23. Kim, J.; Kim, S.; Cho, H.M.; Chang, J.H.; Kim, S.Y. Data sharing policies of journals in life, health, and physical sciences indexed in Journal Citation Reports. *PeerJ* **2020**, *8*, e9924. [[CrossRef](#)] [[PubMed](#)]
24. Tal-Socher, M.; Ziderman, A. Data sharing policies in scholarly publications: Interdisciplinary comparisons. *Prometheus* **2020**, *36*, 116–134. [[CrossRef](#)]
25. Kim, J.; Bai, S.Y. Status and factors associated with the adoption of data sharing policies in Asian journals. *Sci. Ed.* **2022**, *9*, 97–104. [[CrossRef](#)]
26. Aleixandre-Benavent, R.; Moreno-Solano, L.; Ferrer-Sapena, A.; Sánchez-Pérez, E.A. Correlation between impact factor and public availability of published research data in information science and library science journals. *Scientometrics* **2016**, *107*, 1–13. [[CrossRef](#)]
27. González-Pereira, B.; Guerrero-Bote, V.P.; Moya-Anegón, F. A new approach to the metric of journals' scientific prestige: The SJR indicator. *J. Informetr.* **2010**, *4*, 379–391. [[CrossRef](#)]
28. SCImago. Help. Available online: <https://www.scimagojr.com/help.php> (accessed on 13 January 2024).
29. Page, M.J.; Nguyen, P.Y.; Hamilton, D.G.; Haddaway, N.R.; Kanukula, R.; Moher, D.; McKenzie, J.E. Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: A content analysis. *J. Clin. Epidemiol.* **2022**, *147*, 1–10. [[CrossRef](#)]
30. Corti, L.; Bishop, L. Ethical Issues in Data Sharing and Archiving. In *Handbook of Research Ethics and Scientific Integrity*; Springer: Cham, Switzerland, 2020; pp. 403–426. [[CrossRef](#)]
31. DuBois, J.M.; Mozersky, J.; Parsons, M.; Walsh, H.A.; Friedrich, A.; Pienta, A. Exchanging words: Engaging the challenges of sharing qualitative research data. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2206981120. [[CrossRef](#)]
32. White, T.; Blok, E.; Calhoun, V.D. Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Hum. Brain Mapp.* **2022**, *43*, 278–291. [[CrossRef](#)]
33. Sturges, P.; Bamkin, M.; Anders, J.H.; Hubbard, B.; Hussain, A.; Heeley, M. Research data sharing: Developing a stakeholder-driven model for journal policies. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2445–2455. [[CrossRef](#)]
34. Kim, S.Y.; Yi, H.J.; Huh, S. Current and planned adoption of data sharing policies by editors of Korean scholarly journals. *Sci. Ed.* **2019**, *6*, 19–24. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.