Incentivising best practice in research data sharing: Experiments to increase use of and engagement with data repositories

Lauren Cadwallader[1], Beruria Novich[1], Iain Hrynaszkiewicz[1]

Correspondence to: lcadwallader@plos.org

1. PLOS, 1265 Battery Street, Suite200, San Francisco, CA94111, USA

Abstract

Improving the uptake of repositories to share research data is an aim of many publishers, funders and infrastructure providers. Even at the publisher PLOS, which has a mandatory data sharing policy, repositories are still used less commonly than Supporting Information to share data. This preprint presents the results of two experiments that tested solutions that aimed to increase the use of repositories for data sharing as well as increase engagement with shared data. The experiments—integration of the Dryad repository into the manuscript submission system at *PLOS Pathogens* and implementing an Accessible Data icon to signal data shared in a repository on published articles across the PLOS journal portfolio—were designed to be interventions that required minimal extra effort for authors (researchers). We collected usage data on these solutions as well as survey (n=654 and n=4,898) and interview (n=12) data from submitting authors. The results show that author uptake of the integrated repository (used by ~2% of submissions) was lower than expected in part due to lack of awareness despite various communication methods being used. Integration of data repositories into the journal submission process, in the context in which we tested it, may not increase use of repositories without additional visibility, or policy incentives. Our survey results suggest the Accessible Data icon did have some effect on author behaviour, although not in the expected way, as it influenced repository choice for authors who had already planned to use a repository rather than influencing the choice of sharing method. Furthermore, the Accessible Data icon was successful in increasing engagement with shared data, as measured by an increase in average monthly views of datasets linked to a cohort of 543 published articles that displayed it from 2.5 to 3.0 (an increase of 20%) comparing 12-month periods either side of the introduction of the icon. The results of these two experiments provide valuable insights to publishers and other stakeholders about strategies for increasing the use of repositories for sharing research data.

Introduction

Sharing research data alongside published articles is a practice that has been increasing steadily over a number of years (Serghiou et al. 2021, Public Library of Science 2022). However, methods of data sharing vary, as do publisher policies (Hrynaszkiewicz et al. 2020). Best practice in data sharing, following the FAIR principles (Wilkinson et al. 2016), is commonly accepted to mean data are shared in a repository. However, research has shown that, despite numerous policy and other initiatives by journals, publishers and funding agencies designed to increase FAIR data sharing, such as making Data Availability Statements (DAS) compulsory, the proportion of data shared following best practices remains low (Colavizza et al. 2020, McIntosh et al. 2021, McIntosh et al. 2022). Sharing data in a repository brings benefits both to the original authors (researchers) who generate research data and to others who may reuse the data. Using shared data can benefit the researcher by reducing the costs of, or potential for waste in, research; producing larger datasets; and giving researchers access to a broader range of data (Piwowar & Vision 2013, Arza & Fressoli 2017). In turn, this can provide benefits to society, such as through the synthesis of available evidence on health topics (Pisani et al. 2016).

The barriers that prevent researchers from sharing data have been well studied (see Borycz et al. 2023 for a synthesis) and commonly point to reasons including lack of time, uncertainty about which repository to use, prohibitive costs, and concerns about the quality of the data. However, a survey we conducted (Hrynaszkiewicz et al. 2021b) suggested that researchers are on average satisfied with their ability to share data, from their perspective. Given relatively high levels of satisfaction with current data-sharing approaches found in our survey, researchers may lack intrinsic motivation to adopt new solutions, such as data repositories, unless they offer more convenience or other advantages over their current approaches. Conversely, our survey also found that more than half of researchers reuse research data in their own research, but are dissatisfied with their ability to access data for reuse.

PLOS's data availability policy (https://journals.plos.org/plosone/s/data-availability), since 2014, requires all authors publishing in any PLOS journal to make the data supporting the results of their article publicly available (with a limited number of exceptions) and include a DAS in their publications. While data sharing is required under the policy, how data are shared is generally not prescribed by PLOS. The use of data repositories as the preferred method of data sharing is strongly encouraged under the policy, but is not mandatory. This project aimed to increase the proportion of data shared in repositories and the reuse of shared data by testing two solutions that were integrated into existing peer review, publishing, and content discovery workflows, and as such were assumed to be convenient to use, requiring minimal behavioural change for researchers to adopt. We sought to understand what effects the solutions had on data-sharing and reuse behaviours, through usage data, and to understand any effects that were observed, through surveys and interviews. The project was designed in three parts:
1) Integrate a generalist repository with the manuscript submission process at *PLOS Pathogens* to try and increase the proportion of data shared in a repository. *PLOS Pathogens* was selected to trial Dryad (https://datadryad.org/) as the research data its

authors commonly put into Supporting Information files are often not suited to a discipline-specific repository but could be of value to other researchers who may wish to reproduce or utilise their results (Cadwallader et al. 2021). The Dryad repository was chosen because it was readily available for integration via the *PLOS Pathogens* submission system (Aries Systems Corporation's Editorial Manager) and because it can host a wide variety of data types and formats, making it relevant to a high proportion of authors submitting to the journal. Dryad also offers other features to the researcher, such as review of data files by human data curators. Fees associated with depositing data in Dryad were paid by PLOS, as part of our research grant costs, to make the solution free at point of use for authors.

2) Implement an "Accessible Data" icon (badge) on published articles that link to research data in a repository via their DAS. The goals of this icon are to increase readers' engagement with shared data and to offer a novel incentive to authors to share data in a repository. Open data badges on publications have been associated with increased data sharing, by offering a visual reward to researchers on their articles (Kidwell et al. 2016, Rowhani-Farid et al. 2017, Pearce & Grant 2020). As well as signalling that data were shared in a repository, we wanted to provide direct access to the data via the Accessible Data icon. Rapidly-available proxy measures of reuse of data in repositories were chosen: measuring clicks on the icon by readers and, where available, via views and downloads of the data on the repository platform. These proxies were chosen partly for pragmatic reasons as data reuse, when measured through citations of datasets or subsequent publications, occurs over a much longer timescale than our experiments (Lowenberg et al 2019, Piwowar & Vision 2013).

3) Assess these two experiments through surveys and interviews designed to explore the data sharing attitudes and experiences of the researchers submitting to PLOS journals, and offer understanding as to why we observed the results we did. The survey results offer both quantitative and qualitative insights to contextualise and add to the interpretation of the usage data generated by the two experiments.

The experiments were designed to be scalable and repeatable such that similar solutions could be adopted by other journals and publishers. We hypothesised that the Dryad integration would be associated with an absolute increase of 10 percentage points in the use of data repositories by *PLOS Pathogens* authors, from the 2020 baseline of 25% of published articles. We assumed this to be a modest but meaningful increase in use of repositories, based on observing that data repositories integrated with submission systems in other scholarly publishing contexts are used by more than a third of submitting authors (Hufton 2015). Additionally, the high proportion of *PLOS Pathogens* datasets currently being shared as Supporting Information could be submitted to Dryad (74% of articles use SI, some in combination with a repository). We also hypothesised that the Accessible Data icon would be associated with a statistically significant increase in engagement with datasets linked from articles with the feature, and that the icon would motivate authors to deposit their data in a repository in order to have the icon applied to their article.

Methods

Dryad integration and survey

The Dryad data repository was integrated with the *PLOS Pathogens* submission system in Editorial Manager using the mechanism provided by Aries Systems Corporation, which operates Editorial Manager (Cadwallader et al. 2021). Editorial Manager allows for limited customisation of the integration through short descriptive text shown to the user when submitting their manuscript to the journal. The option to use the Dryad integration was listed under the heading "Research Data" in the list of content to upload along with the user's manuscript file. The integration was enabled for authors submitting from the 5th of October 2021. After six months, user testing was conducted to refine the presentation of the integration in the Editorial Manager system with the aim of improving uptake. Edits were also made to the help text related to the integration and the layout of the information. The user testing confirmed that "Research Data" was an appropriate title under which to provide a link to the integration.

Information about the Dryad integration was added to the submission guidelines and webpages about data availability for *PLOS Pathogens* as well as a general page on Open Science on the PLOS website. Information about the integration was also added to standard letters sent to authors during the submission and peer-review process and the integration was written about in an editorial published in the journal (Cadwallader et al. 2021). Social media marketing and banner ads on *PLOS Pathogens* pages linking to a blog on the integration also took place, as well as email marketing communications sent to researchers in the pathogens research community.

Use of the Dryad integration was monitored via the administrator interface on the Dryad website. This gave information on the data deposited in Dryad and the manuscript number of the related submission in Editorial Manager. Manuscript number was used to identify instances of multiple deposits to Dryad by the same author for the same manuscript. A survey related to the Dryad integration was sent to all corresponding authors submitting a research article manuscript to *PLOS Pathogens* between the 5th of October 2021 and the 23rd of July 2022, with reminders to participate sent out 6 and 10 days after the initial invitation. An incentive was offered of a US$200 gift card for 3 respondents.

Accessible Data feature and survey

The Accessible Data icon was designed to display on the html version of published PLOS articles that linked to research data, via the DAS, shared in one of three generalist repositories: Figshare, Dryad, and the Open Science Framework (OSF). The icon is displayed automatically on articles that meet these criteria, on PLOS's publishing platform. Automation was an important design choice with the icon to ensure the approach was low cost and could scale easily across thousands of published articles. This approach differs from other Open Science badging approaches, which usually require human intervention—by authors and/or journal staff—to

implement (Grant & Pearce 2020). The Accessible Data icon was introduced across all PLOS journals on the 29th of March 2022 and was initially included on 4,892 articles. After 12 months a further 1,212 newly published articles, which met the criteria for displaying the icon, also featured the icon. The icon includes a "See the Data" button (Figure 1), which links to the data in the repository, allowing readers to go from the article to the data in one click.
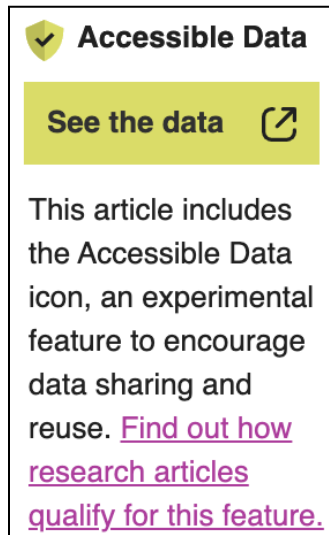
Figure 1: The Accessible Data icon

A survey related to the Accessible Data icon was sent to corresponding authors submitting a research article at seven PLOS journals the week after their submission with reminders to complete the survey sent 6 and 10 days later. Surveys were sent in the period from the 20th of April 2022 to the 15th of August 2022. The survey sent during this time period to authors submitting to *PLOS Pathogens* also included the questions related to the Dryad integration so that these authors were presented with one survey. No incentive was offered to authors from the other six journals to complete the survey.

Survey analysis

For each of these two surveys–Dryad integration and Accessible data–response frequencies were run for all respondents and for key respondent segments. For all questions, differences in the responses of respondent segments were tested for statistical significance at the 95% confidence level. Only segments with 30 or more respondents were tested for significance. Respondents were segmented by author demographics (e.g. career stage, region), data sharing behaviours, awareness of the feature and journal/sub-discipline. All survey analysis was conducted using Uncle Group data tabulation software. Statistical tests of the respondent segments used the software's SGTEST for performing significance testing between column percentages using a t-distribution.

Interview recruitment

Interview invitations were emailed to researchers who had interacted with the PLOS Accessible Data icon as either:
- an author of a PLOS published article which had shared data in Dryad, Figshare, or OSF and had the Accessible Data icon applied to their article
- or an individual who viewed data shared through the Accessible Data icon in a PLOS article and provided their email when prompted to take part in research via a pop-up.

Invitations were sent to 692 authors and 91 to readers of articles. Those who agreed to participate completed a screener where they indicated their experience as well as availability for participation and were provided with information for participants. The screener was completed by 109 authors and 12 readers. Final participants (9 authors and 3 readers) were selected based on the following priorities: Recent experience submitting to a PLOS journal or clicking on the Accessible Data icon was favoured; diversity in repository used (Dryad, Figshare, OSF); and broad range of journals among participants (*PLOS ONE*, *PLOS Biology*, *PLOS Computational Biology*, *PLOS Climate*, *PLOS Genetics*).

Interviews were conducted via Zoom using audio and video (with the exception of one participant who preferred not to use video), recordings were used for transcriptions which were outsourced to Rev transcription services to ensure transcription accuracy. Interviews followed a semi-structured approach—an interview guide was used to ensure key questions were covered, but participants were encouraged to talk through their process in their own words, and some questions were covered in the context of other responses. After the interviews, quotes relevant to the research questions were selected, collated and coded.

Ethical considerations

Approval from a research ethics committee was not sought for this work as we considered the research to be low risk. Sensitive information about the participants was not collected during the surveys or when gathering usage data from Dryad and Figshare and all data were collected anonymously. Survey participants were informed that participation was voluntary, and that they were free to withdraw at any time until they submitted their response. The survey results and data from the platforms (Figshare, Dryad, PLOS) were only analysed in aggregate and answers were never associated with individual participants. Interview participants were informed of the nature and purpose of the study and were asked for consent to participate. They were free to withdraw consent at any time. The data collection procedures and survey tool are compliant with the General Data Protection Regulation 2016/679.

Results

The data resulting from this work is shared in Figshare in a Collection (Cadwallader & Novich 2023). The Collection includes: the survey results, usage data and interview materials. The interview transcripts have not been shared, to protect the participants' identity.

Dryad integration results

Usage data results

Use of the Dryad integration was monitored for the 12 months after it was introduced. Information recorded included the manuscript number that the user accessed the integration via, how many completed deposits, the overall outcome of the manuscript after evaluation by the journal and deposit of the associated dataset.

In total, 44 authors of manuscripts (2.1%) submitted to *PLOS Pathogens* engaged with the integration during the first 12 months, i.e. clicked through from the Editorial Manager system to the Dryad system. It is possible for an author to click through to the Dryad system from Editorial Manager multiple times and each of these events is logged even if no action takes place within the Dryad system. Sixteen of the 44 authors clicked through to Dryad multiple times with the most common number of clicks being 2 and the maximum number of clicks being 12. Dryad deposits were completed for 28 out of the 44 associated manuscript submissions, although 4 manuscript submissions were never completed. Eleven of the authors who completed Dryad deposits had clicked through to the Dryad system more than once.

Twelve manuscripts associated with completed Dryad deposits were ultimately accepted for publication and half of these retained the Dryad deposit information in the published article. For the six that removed the Dryad deposit from the final article, five chose to share data via Supporting Information instead and one via a discipline-specific repository, however, for two of these articles, the data shared at publication was not an exact replica of the data previously uploaded to Dryad. 73% (n=32) of authors who engaged with the Dryad integration did so at the time of their initial manuscript submission to PLOS, although 4 of these did not complete the manuscript submission process. The remaining 27% engaged with the integration at the point where they had been asked to revise their manuscript during the editorial process. Two authors who had used the integration on initial submission also engaged with it again at revision via a new deposit.

Survey results

The survey was sent to a total of 1,703 unique authors and 654 completed the survey (38.4% response rate) by answering all the required questions. The survey asked about the authors' data sharing behaviours both for the article submitted to *PLOS Pathogens* and for previous article submission; their reasons for choosing the methods of data sharing they did; and their awareness and understanding of the Dryad integration (Novich et al. 2023).

Demographics and past sharing behaviours

Of the respondents sent the survey for *PLOS Pathogens* authors (n=654), 75% identified as principal Investigators, with the remainder of respondents being post-doctoral researchers (9%), staff scientists (8%), graduate students (5%) and other (3%). Half of respondents identified themselves as mid-career (51%), with the remaining split approximately equally between early-

(24%) and late-career (25%). 77% of respondents had more than 10 years of research experience, including 40% with 21 or more years experience. The top four regions represented by respondents were Europe (28%), North America (26%), China (25%) and Asia-Pacific (excluding China, 16%). Virology was the most represented *PLOS Pathogens* section (48%), followed by Bacteriology (26%) and Parasitology (13%). The remaining sections–Plant Pathogens, Mycology, Genomics, Evolutionary Biology and Prions–contributed between 1-5% of the sample each.

Respondents were asked how they most often shared data for manuscripts they had previously submitted. Just over half (51%) had shared data in Supporting Information or within the manuscript. Other common responses were in a repository (22%), and did not previously submit any manuscripts with shared data (20%). A small number of respondents said they most often shared data on request (6%).
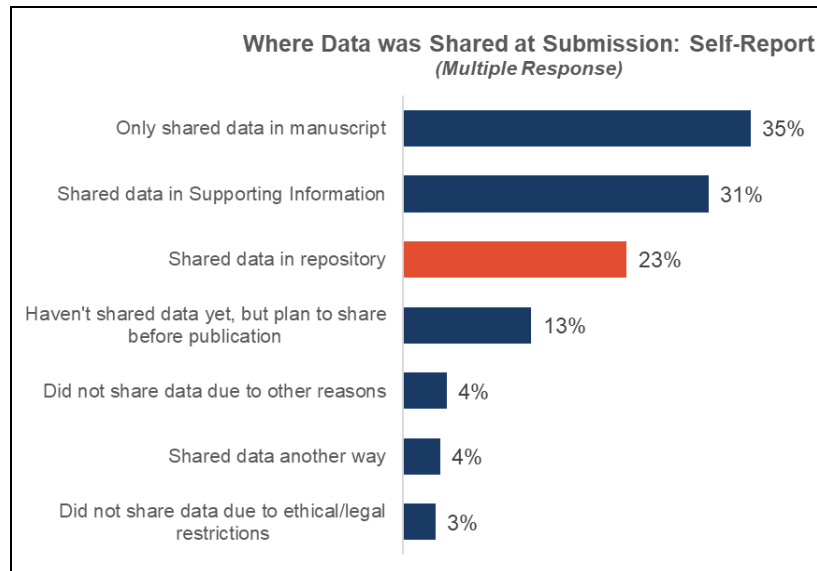
Awareness

Respondents (n=654) were asked when they became aware of the Dryad integration, if at all, and how they learned about it. Only 16% were aware of the integration before submitting their manuscript, with a further 37% becoming aware during the submission process. Almost half (47%) were unaware of the Dryad integration before taking the survey. The journal website was the main source of information for those who were aware before submission (64%), followed by emails from PLOS (31%). Of those who were aware of the feature (n=344), 51% felt they understood how to use the integration.

Data sharing practices

Respondents (n=654) were asked how they shared the data associated with their most recent manuscript submission. The majority of respondents had shared their data in either the manuscript (35%) or Supporting Information (31%) and just under a quarter had used a repository (23%) (see Figure 2). The 23% that used a repository included 8 authors who had used the Dryad integration to deposit data. A number of respondents had not shared their data at the time of submission, with 13% stating that they planned to share their data before publication, 3% had not shared due to ethical/legal restrictions and a further 8% had not shared for other reasons.

Where Data was Shared at Submission: Self-Report
(Multiple Response)

Only shared data in manuscript	35%
Shared data in Supporting Information	31%
Shared data in repository	23%
Haven't shared data yet, but plan to share before publication	13%
Did not share data due to other reasons	4%
Shared data another way	4%
Did not share data due to ethical/legal restrictions	3%

Figure 2. Where was data shared at the point of submission? Data shared in a repository includes data shared via the Dryad integration. Results from the survey of PLOS Pathogens *authors (n=654).*

Those who hadn't shared yet and were planning to share before publication (n=85) were asked where they intended to share their data and were able to select multiple options. Answers were evenly spread across the options of data shared in the manuscript (34%), Supporting Information (35%), via the Dryad integration (34%) and in a repository not via the integration (34%). Four authors (5%) planned to share data in another way.

Reason for choosing the Dryad integration (or not)

Respondents were asked to select the reasons why they did or did not choose to use the Dryad integration, with multiple responses permitted. The main reasons for using the integration as given by respondents who used or planned to use the integration (n=37) were: to comply with PLOS' data policy (49%), to make it easier for others to find the data (46%), to make it easier for reviewers and editors to access the data (43%), to make the research open (41%) and the author was under the impression that the Dryad repository was recommended by PLOS (38%). Other reasons for using the Dryad integration scored less than 20%, including: it was easy to upload the data; it was free to use; and it was available during manuscript submissions. Responses that scored less than 10% included: the author wanted to take advantage of the curation checks offered by Dryad; and the author wanted to receive credit for sharing their data.

For those who did not share or plan to share data using the Dryad integration (n=574) the biggest reason for not using it was lack of awareness (35%), followed by uncertainty in how to use the integration (26%), the data had already been uploaded to another repository (21%), the author had concerns about others using their data before they had used it for additional publications (11%), lack of time (9%), the data required a different repository (7%) and it was too much effort to share data in a repository (7%). Other, less influential reasons given by less than

7% of respondents were concern about applying a CC0 waiver to their data, concerns about other researchers not understanding the data, and concerns about the Dryad curation checks.

The survey asked respondents (n=654) what impact the Dryad integration would have on their likelihood to share data in a repository in a future submission. A positive impact (more likely or much more likely) was recorded for 44% of respondents. A negative impact was reported by only 12% of the respondents.

Eight respondents had used the Dryad integration during submission and rated their satisfaction with specific aspects of the integration. Six of the eight respondents were satisfied (satisfied or very satisfied) with the overall process of using the integration. All eight of the respondents were satisfied with the clarity of instructions and information provided during the upload process, five were satisfied with the information they needed to provide, seven were satisfied with the confirmation that the upload was successful and six were satisfied with the time it took to upload .

Accessible Data icon

Usage data results

Engagement with the icon on the PLOS platform

The Accessible Data icon was clicked on at least once for the vast majority of articles that displayed it–86% of articles (n=4,892) with the icon added at launch and 99% of articles (n=1212) that were published with the icon after the initial launch. The articles with the icon at launch that received at least one click accrued, on average, 3.8 clicks on the icon during the 12 months following the launch. Articles that had the icon added post-launch (i.e. articles published between the 30th of March 2022 and the 31st of March 2023) and received at least one click accrued, on average, 4.0 clicks on the icon despite being online for only a portion of the study period.

For the articles with the icon at launch, the average number of clicks over the 12 month period ranged from 3.2 (+/- 8.3) to 5.2 (+/- 15.7) when divided into publication years (2014-2022). The highest value, 5.2, belongs to the 2014 publications, which has the smallest sample size of all the years. Publications from 2021 and 2022 have slightly higher average click rates (4.7 and 4.8 respectively) than the rates from 2015-2020 (see Table 1).

Table 1: Data for the articles featuring the Accessible Data icon from launch.

Publication year	2014	2015	2016	2017	2018	2019	2020	2021	2022	Overall
Number of articles	46	427	435	482	560	612	624	797	223	4206

Average number of clicks on AD icon	5.2	3.3	3.3	3.4	3.2	3.5	3.9	4.7	4.8	3.8
Stdev	15.7	8.3	5.7	6.3	4.9	7.7	6.9	7.4	6.8	6.8
Min number of clicks	1	1	1	1	1	1	1	1	1	1
Max number of clicks	105	143	67	73	44	164	79	80	86	164
Average number of article views	1094.4	914.2	473.3	543.9	4160.0	781.7	872.8	1364.1	1696.8	1361.6

Engagement with linked datasets deposited in Figshare

Views and downloads data are publicly available for datasets deposited in Figshare, one of the 3 repositories that was supported by the icon. Different repositories record views and downloads in different ways (Lowenberg et al. 2019), limiting the ability to compare the effect of the icon between repositories. Views and downloads data for a cohort of 543 Figshare records that were linked to from a PLOS article with the Accessible Data feature were used to explore potential correlations between the icon and views and downloads, with support from Figshare. All of the 543 articles in this cohort had the feature added at the point of launch. Monthly views and download figures were analysed from March 2021 to April 2023 and bot activity was removed so that only views by human actors were counted. These articles originated from seven PLOS journals (see Table 2).

Table 2. Number of articles by journal with Figshare links that are included in the Figshare views and downloads data.

Journal	n articles included in Figshare cohort (n= 543)	% of sample
PLOS ONE	490	90%
PLOS Computational Biology	18	3%
PLOS Biology	11	2%
PLOS Neglected Tropical Diseases	11	2%
PLOS Genetics	7	1%
PLOS Pathogens	4	1%
PLOS Medicine	2	<1%

The articles included in the Figshare views and downloads cohort were published between 2014 and 2022. *PLOS ONE* was the only journal with articles in each year. Focusing on the data from 12 months on either side of the feature's launch (April 2021 to March 2023), the number of views received in any given month ranges from zero to 340. The average number of views received per month was 2.5 (+/- 9.8) in the 12 months prior to the launch of the feature and 3.0 (+/- 10.3) in the 12 months following the launch, a 20% increase in views compared to the pre-launch period. There are several datasets with very high numbers of views per month (i.e. an order of magnitude greater than the mean (or more)) as well as many datasets that received zero views. Comparing the pre- and post-launch data, the number of datasets with zero monthly views decreases and the number of datasets with 1-10 views per month (on average) increases (Figure 3). The number of downloads ranges from zero to 193. The average number of downloads received per month was 0.4 (+/- 3.2) in the 12 months prior to the launch of the feature and 0.6 (+/- 2.4) in the 12 months following the launch.
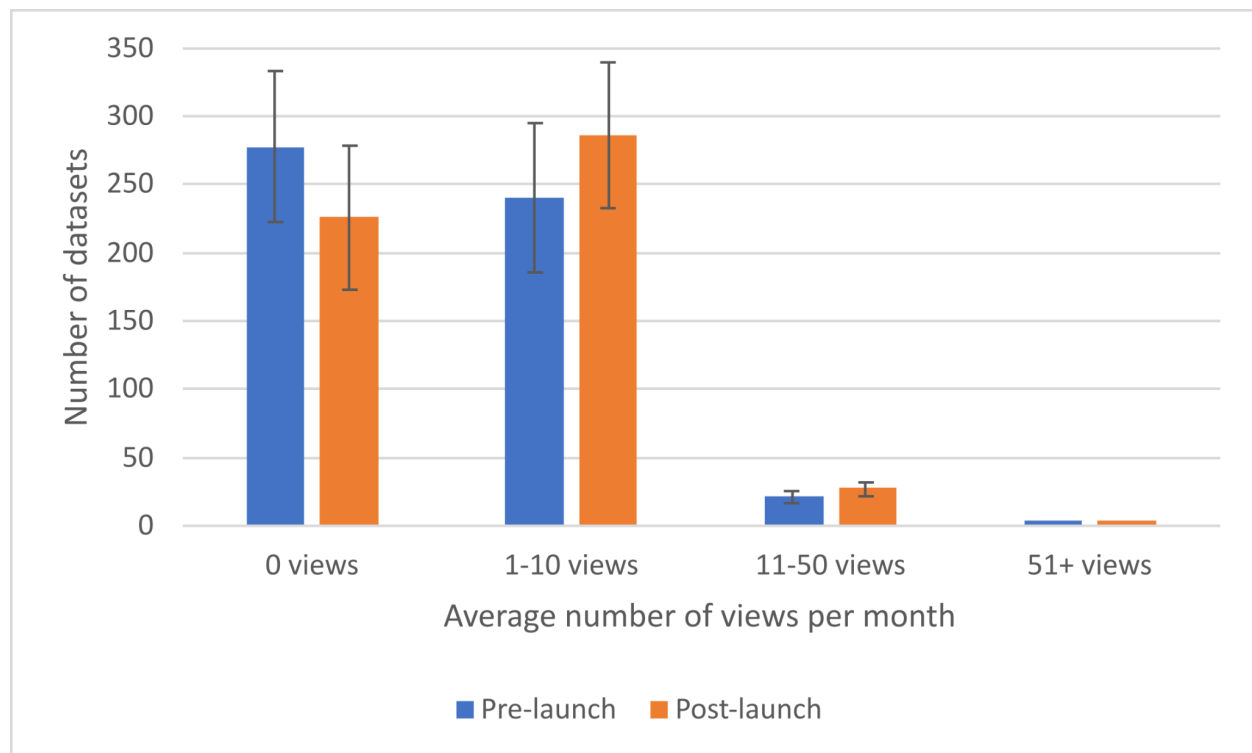
Figure 3: Average dataset views per month for 12 months pre-launch (April 2021-March 2022) and 12 months post-launch (April 2022-March 2023). **Total number of datasets in each month equals 543. The standard deviation for the 51+ views is +/-1 and +/-0 for pre- and post-launch respectively.**

As the data is zero-inflated and overdispersed, a zero-inflated negative binomial model has been applied to the view data, which shows that the increase in views from an average of 2.5 pre-launch to 3.0 (20% increase) in the post-launch period is statistically significant but a small difference. The increase in the number of datasets with 1-10 views per month in the post-launch period is not statistically significant.

Survey and interview results

The Accessible Data survey was sent to a total of 11,549 authors and was completed by 4,670 authors (39.4% response rate). The combined Dryad/Accessible Data survey was sent to an additional 566 authors submitting to *PLOS Pathogens* and was completed by 228 authors (40.3% response rate). Therefore the final Accessible Data survey dataset contains 4,898 complete responses.

Demographics and past sharing behaviours

Survey and interview data were gathered in two phases—a survey related to the Accessible Data icon was sent to all submitting authors at seven PLOS journals and follow-up interviews were conducted with 12 participants.

The Accessible Data survey received 4,898 responses. Half (50%) of the respondents identified as Principal Investigators, 16% as graduate students, 16% as staff scientists, 8% as post-doctoral researchers, and 9% as other. Respondents were also asked their career stage, with 36% identifying as early career, 46% as mid-career and 18% as late career. They were also asked how many years they had been an active researcher. Over half (54%) have been an active researcher for up to 10 years, 27% for 11-20 years, and the remaining 20% for 21 or more years. The top five regions represented among respondents were Europe (23%), Research 4 Life countries (21%), China (18%), Asia-Pacific (excluding China, 17%) and North America (14%). The majority of respondents had published in *PLOS ONE* (80%), with *PLOS Pathogens* contributing 5% of respondents. The remaining journals (*PLOS Computational Biology*, *PLOS Neglected Tropical Diseases*, *PLOS Biology*, *PLOS Medicine* and *PLOS Genetics*) each contributed 3% of respondents.

Respondents (n=4,898) were asked how they most often shared data for manuscripts they had previously submitted. 40% most often shared data in Supporting Information or within the manuscript, 29% had not previously submitted a manuscript with shared data, 16% most often shared in a repository, 13% shared privately on request and 1% had most often shared in another way.

Awareness

A third (34%) of respondents (n=4,898) were aware of the Accessible Data before submitting to PLOS. A further third of the sample (35%) became aware of the feature during submission and 31% remained unaware of the feature before completing the survey. Journal web pages (62%), existing articles displaying the feature (34%) and emails from PLOS (20%) were the most common sources of information about the feature prior to submission.

Data sharing practices

Respondents (n=4,897) were asked to indicate where they had shared their data at the point of submitting their article to PLOS. The majority (57%) had shared in either the manuscript or Supporting Information, 15% had shared in a repository, 13% had not shared due to ethical or legal restrictions, 12% had not shared but planned to and the remaining 9% had either shared in

a different way or had different reasons for not sharing. Use of a repository to share data at the point of a submission was statistically significantly higher for four PLOS journals than for the other three (see Table 3).

Table 3. Self-reported repository usage at the point of submission segmented by journal. Superscript letters denote the levels of statistical significance - Journals with [a] are higher than all others and journals with [b] are higher than journals with [c]. Data from the Accessible Data survey.

Journal	Submissions sharing data in a repository (n)
PLOS Computational Biology[a]	51% (85)
PLOS Biology[b]	33% (54)
PLOS Genetics[b]	33% (40)
PLOS Pathogens[b]	29% (66)
PLOS Neglected Tropical Diseases[c]	13% (21)
PLOS ONE[c]	12% (477)
PLOS Medicine[c]	10% (15)

Post-doctoral researchers were statistically significantly more likely to have shared data in a repository, as were authors from the US/UK/EU compared to all other countries.

Of the respondents who had already shared or planned to share data via a repository (n=1,010), 30% used (or planned to use) one of the three repositories eligible for the Accessible Data icon (Dryad, Figshare, or OSF), 53% used (or planned to use) another repository and the remaining 17% were not sure which repository.

Respondents who were aware of the Accessible Data feature and who had used a repository (n=643) were asked to what extent the feature influenced their decision to share their data in a repository. 51% answered that they were influenced (either somewhat or strongly) by the existence of the feature and 49% answered that they were not influenced at all.

Respondents (n=4,898) were asked to think about their potential future submissions to PLOS journals, and what the impact they feel the availability of the Accessible Data feature will have on their likelihood to share data in a repository. 40% thought they would be more or much more likely to use a repository, 29% thought that it would have no impact, 10% thought the feature would make them less likely to use a repository and 21% were not sure what impact it would have.

Participant interviews

The 12 semi-structured interviews covered topics related to data sharing. The reasons the interviewees gave for sharing data in a repository could be classified into a few themes: journal or funder requirements, expectations of the research culture, and technical reasons, e.g. data preservation or enhancing discoverability. The time taken to complete data deposits was seen as a negative by four of the interviewees. This feeling was often compounded by the scarce attention paid to data by others and the lack of credit depositors expected to receive.

On the topic of data discoverability, the interviewees generally felt that they knew where to find (a link to) shared data in a research article. In relation to the Accessible Data icon itself, there were mixed responses with three interviewees commenting very positively on the icon, two commenting positively on its usefulness, and one taking a more critical view as the icon could confer unwarranted credibility on the article.

Discussion

The results from these experiments provide a rich dataset lending insight into behaviours, attitudes and practices around data sharing in the contexts of our experiments as well as more broadly, given the large number of survey responses we received (survey responses are combined for some questions). First we look at some key observations related to the individual experiments and some common themes that surfaced in both experiments, before discussing the effectiveness of the experiments in the context of the original research questions. We conclude by reflecting on the potential implications of our findings for different stakeholders (publishers, funders, infrastructure providers etc).

Repository use and other data sharing behaviours

When asked about how they have most often shared data in the past (n=4,898), only 56% of Accessible Data survey respondents had shared data (across 7 PLOS journals), with the majority (40% of respondents) sharing data most often in Supporting Information or the manuscript and 16% sharing most often in a repository. The remaining 44% had either not shared data or only done so on request. *PLOS Pathogens* specifically has a higher rate of submitting authors with data sharing experience (73%) compared to the average across all journals although Supporting Information or within the manuscript are the most popular avenues for this. While this suggests that there is a high number of authors to potentially 'convert' to using a repository for data sharing, it also implies that there is a large proportion of authors who are, upon submission to the journal, potentially unaware of how or why they should share data because of their previous experience (or lack thereof), who are then expected to share data to comply with PLOS' mandatory data availability policy (https://journals.plos.org/plosone/s/data-availability).

Of all respondents, 15% shared their data in a repository on submission to PLOS, which is a similar proportion to the number of authors who most often shared data in a repository in the

past (16%). On submission, 56% shared data in Supporting Information or the manuscript, which is higher than the proportion who reported that this method of sharing is the one they practised most often in the past (40%). For both past and present behaviour, the most popular method of data sharing (using Supporting Information or the manuscript) is the easiest from the authors' perspective, but suboptimal from a data reuse perspective. From the survey results we can also see that current behaviours are strongly linked to previous behaviour, with those using a repository in the past more likely to do so when they submit to PLOS (52% for previously most often repository compared to 9% or less for others) and those sharing in less optimal ways (or not sharing at all) more likely to share data in the SI or manuscript (81% for previously most often in Supporting Information or manuscript compared to 30% for previously most often in a repository).

Among submitting authors to PLOS journals, 18% said they hadn't shared data at the point of submission (they either plan to share before publication or had stated they did not share data due to ethical/legal restrictions or other reasons). As PLOS has a mandatory data sharing policy, this is a sizeable proportion of authors who could be influenced in their choice of data sharing location, assuming that the accept rate is equal across all categories of data sharing on submission. Of those planning to share at a later point (n=596), the proportion saying they plan to share via Supporting Information or the manuscript (59%) is comparable to the authors who shared data using this method on submission (57%), however, 51% of those planning to share state they will use a repository (either exclusively or in addition to other sharing methods), which is much higher than the rate reported by those who shared data in a repository on submission (15%) and also higher than the observed rates of repository use in published PLOS articles (28% shared in a repository in 2022) (Public Library of Science 2022). From these different data sources, it appears that there is a segment of authors who are unwilling or unable to invest the time in sharing data in a repository upon submission but many have good intentions for sharing data and some follow through on them.

There are small but statistically significant differences in where data are shared, both on submission to PLOS and in previous submissions, between researchers at different career stages. Early-career researchers are slightly more likely to share data in a repository both on submission to PLOS and for previous research articles than middle- and late-career researchers (previous repository use 18%/15%/14% and repository use on submission to PLOS 17%/14%/16% for early/middle/late career). Early-career researchers have also been shown to be more likely to practise open science behaviours in other studies (Tenopir et al. 2011, Campbell et al. 2019, Toribio-Flórez et al. 2020, Hrynaszkiewicz et al. 2021a).
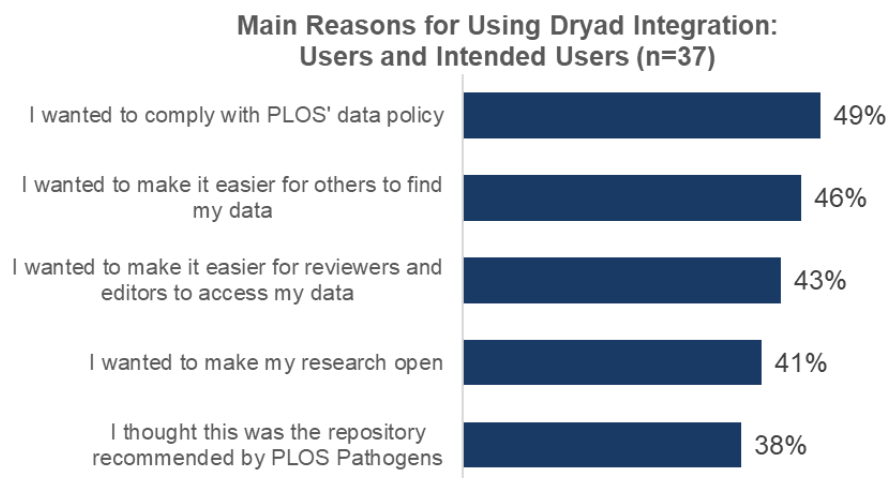
Dryad integration

Previous research (Hrynaszkiewicz et al. 2021b) has shown that researchers are satisfied with their data sharing practices and are unlikely to seek out new data sharing solutions. The Dryad integration was implemented and was free at the point of use with an assumption that it represented a small change to an existing workflow and therefore would not require a significant change in author behaviour to be used. The multiple times that Dryad submissions were started

but not completed, evident in the usage data, indicates that some authors were at least willing to engage with the repository but were unsure whether or not to use it or needed more information to decide if it was the correct choice for them to share their data. This speaks to a general unfamiliarity with data sharing requirements of repositories or with Dryad specifically. Dryad has some additional requirements compared to sharing data in other generalist repositories, such as Figshare, Zenodo, Open Science Framework, including checks by data curators, which can help to improve data quality and safeguard against sharing of sensitive datasets. However, the reasons given for not using Dryad we observed in our survey, aside from awareness of the Editorial Manager integration, are similar to the reasons found by other studies (Van den Eynden et al. 2016; Tenopir et al. 2020; Hrynaszkiewicz et al. 2021b; Gomes et al. 2022) and they do not suggest authors' concerns about using a repository were specific to Dryad.

The biggest motivators for using the Dryad integration to share data related to support for or recognition of benefits of open science, and related to policy compliance (Figure 4). These findings align with those of other studies, e.g. Digital Science (2022) found compliance and public benefit to be equal motivators for data sharing. Authors were less motivated by the offerings unique to the integrated repository solution, such as convenience in the workflow or the added curation checks available from Dryad. However, neither did authors appear to be put off from using the integration by the features it offered.

Main Reasons for Using Dryad Integration:
Users and Intended Users (n=37)

I wanted to comply with PLOS' data policy	49%
I wanted to make it easier for others to find my data	46%
I wanted to make it easier for reviewers and editors to access my data	43%
I wanted to make my research open	41%
I thought this was the repository recommended by PLOS Pathogens	38%

Figure 4: Top five reasons why users chose to use the Dryad repository integration on **PLOS Pathogens**. *Other reasons given were less influential, cited by <20% of respondents.*

The use of Dryad by authors on initial submission versus on submission of a revised manuscript after peer review, as well as the number of authors (13%) choosing to delay data sharing until their article reaches a later publication stage (e.g. acceptance), suggests that a repository integrated with the manuscript submissions process may be used by more authors if attention is focused on authors using it closer to the point of acceptance for publication. Integrating with journal submission systems has been seen as a key mechanism for increasing sharing of preprints (Ni & Waltman 2023) and presenting a suite of integrations to facilitate the adoption of several open research practices during submission or revision may be an interesting avenue to

explore in the future. However, further research—such as with ethnographic methods—is needed to more fully understand why researchers do or do not opt in to using optional solutions that support open science practices when presented with them during the manuscript submission process.

Accessible Data icon

Awareness of the icon did not have any impact on the self-reported use of repositories for submitting authors, and in fact, survey responses showed that those aware of it were statistically slightly less likely to have shared their data in a repository. However, the icon only displayed on articles that linked to three generalist data repositories, which may not have been suitable for some researchers. The three repositories that the icon supported in its first iteration (Dryad, Figshare, OSF) were used by 31% of PLOS articles that shared data in a repository in 2021 (Public Library of Science 2022). Discipline-specific repositories, such as Gene Expression Omnibus and Sequence Read Archive, are the repositories of choice for researchers working with certain types of biological data, for example.

Those aware of the Accessible Data icon were statistically significantly more likely to rate four data repositories features as more important than respondents who were unaware, although the difference was relatively small (4-6%). These four features were more specific to the features offered by the Accessible Data icon and Dryad integration solutions and it may be that awareness of these features and their benefits helped to make them more important to respondents (Table 4).

Table 4: Percentage of respondents who rated each feature as very or fairly important segmented by the respondents' awareness of the Accessible Data icon. The "Significant?" column indicates if there was a statistically significant difference between the two cohorts.

Feature	Aware	Not aware	Significant?
Sharing my data to support open science	68%	69%	No
Sharing my data to increase the reproducibility of my work	68%	67%	No
Sharing my data in a way that maximises its potential for discovery and reuse	66%	65%	No
Having access to repository services available to me for free	66%	66%	No
Having my data preserved for the long term	63%	62%	No
Being able to comply with funder and institution requirements about sharing data	63%	63%	No
Being able to access help and support from the publisher to enable me to share data	60%	54%	**Yes**

Having a prominent link to the data from the article page	59%	55%	**Yes**
Being able to submit my manuscript and share my data as part of the same process	58%	53%	**Yes**
Having my data curated by the repository (e.g. checking that data is anonymized, has a clear title, etc.)	57%	53%	**Yes**
Getting credit for sharing my data as a research output in its own right	56%	54%	No

Both the survey and interview data suggest that the Accessible Data feature may have some impact on the future behaviours of researchers when it comes to sharing data in a repository. 45% of authors who most commonly used Supporting Information to share research data previously said they were more likely to share data via a repository in the future (regardless of their awareness of the Accessible Data icon)—the highest percentage of any group based on previous sharing behaviour. A few of the interview participants felt that the icon made shared data more prominent which could help to normalise the practice of sharing data. Therefore, while the initial 1-year experiment with the icon was likely too short for it to have had much effect on data sharing behaviours that can be detected in published articles, it could contribute to the growing normalisation and trend of increasing repository use (Public Library of Science 2022) in the future.

Engagement with data shared in repositories

Usage data

Readers clicked on the Accessible Data icon on articles from all years included in this study (2014-2022). Of those articles, 5,406 received at least one click on the icon out of a total of 6,104 articles that displayed the icon in the study period. This suggests that there is interest in data accompanying research articles across the disciplinary spectrum that PLOS publishes and that engagement with data is not restricted to certain fields or types of data.

The average number of times an article has to be viewed for one click on the Accessible Data icon to take place varies between journals but it is difficult to draw insights due to the small sample sizes for some journals. If the multidisciplinary journal *PLOS ONE* is used as a benchmark of 342.8 article views per click, the data suggest that readers in most of the subject-specific PLOS journals engaged with the icon more readily (see Table 5) although this line of investigation warrants a larger dataset.

Table 5: Average number of article views per click on the Accessible Data icon per journal.

Journal	Average number	Number of

	of article views per click on Accessible Data icon	articles with icon
PLOS Digital Health	109.4	13
PLOS Neglected Tropical Diseases	159.8	65
PLOS Global Public Health	163.8	50
PLOS Computational Biology	164.7	258
PLOS Genetics	175.6	94
PLOS Climate	189.2	9
PLOS Pathogens	202.7	29
PLOS Water	212.4	7
PLOS Biology	231.4	178
PLOS ONE	342.8	4,569
PLOS Medicine	344.0	48
PLOS Sustainability and Transformation	433.9	2

For the subset of articles where dataset views and downloads are available from Figshare, an increase in dataset views was observed, compared to the year before the introduction of the icon, regardless of the year that the dataset was deposited and the associated article published. Previous studies measuring engagement through citations have found it common for datasets to be used for at least 6 years after being made available (Chow et al. 2023; Piwowar & Vision 2013), so it is unsurprising that these datasets in Figshare continue to be of interest even if we do not know if they will be cited in future studies. The increase on average in views of the 543 Figshare datasets after the addition of the icon by 20% indicates that the icon may facilitate discovery of research data. This contrasts with the belief on the part of some interviewees that readers would find the link to the data in the article regardless of where it was placed. The usage data suggests that the Accessible Data icon may give prominence to the data that existing Data Availability Statements do not.

The Accessible Data icon appears on a subset of publications which link to data shared in repositories via their DAS, and this practice has been correlated with increased citation rates compared to similar articles that share data in other ways (Colavizza et al. 2020). Further, data sharing has been linked with perceptions of trust (McIntosh et al. 2022). Having established a

baseline for reader engagement with data sharing in generalist repositories in this experiment, further experimentation could explore any differences in engagement between different research topics/ disciplines, types of data linked to, or type of repository. Kim (2021) found that attitudes towards data reuse by STEM scientists are largely influenced by repository availability (as well as metadata standards and data re-use norms). The reader may not have known that only articles shared in a repository were eligible for the icon but the increase in downloads seen in the Figshare dataset may well support these previous findings. Notions around trust and reuse perhaps come together when data is presented in a prominent way such as with the Accessible Data icon.

Our interview questions focusing on data reuse–asked of the three interviewees who were readers of PLOS articles with the Accessible data icon–revealed that participants did not reuse data to repeat experiments nor did they expect others to do the same with their shared data. Instead, the most common use of shared data was to check the data related to a particularly interesting article, although in some cases just the presence of shared data was sufficient to lend credibility to the article's findings. Other reasons given by the interviewees for accessing shared data was to check the type of data collected by studies similar to theirs to help refine their own methodology.

How effective were these experiments at incentivising use of data repositories and data reuse, and why?

One reason the Dryad integration did not achieve our expected level of use, based on the survey data, likely relates to the perceived lack of information about the option to use Dryad. Only 39% of authors submitting to *PLOS Pathogens* were satisfied or very satisfied with the information provided whereas 81% were satisfied or very satisfied with the submission process overall. Information was provided in submission guidelines, the submissions system (Editorial Manager) and in letters sent to authors. This perhaps indicates that the presentation of the information could be improved, for example, by ensuring it is more prominent rather than part of a long body of text. Of those submitting authors who used or planned to use the Dryad integration, 49% (n=18) did so because they wanted to comply with the PLOS data sharing requirements and 38% (n=14) thought Dryad was the repository recommended by PLOS (Dryad is a repository recommended by PLOS, but not exclusively). Other prominent reasons for using the integration were around making the data open and accessible. However, the low number of actual users of the integration (n=8) and associated survey responses makes it difficult to draw firm conclusions about why it was used, from the data we have.

The survey data on perceptions of the Accessible Data icon and data sharing behaviours showed that the icon may have an effect on the choice of repository for those already prepared to use one. Among authors already sharing in a repository, those aware of the feature were statistically significantly more likely to share in the three repositories included in the feature (Dryad, Figshare, or OSF) than those not aware (37% v 21%). This suggests that behaviours of those who are already practising Open Science can be modified with this kind of solution but not to the extent that it changes a researchers' overall data sharing behaviour. However, when

asked about future likelihood of using a repository, the survey results suggest that behaviour might change, with 40% of all respondents saying they were more likely to use a repository. The response was statistically significantly higher (45% answered "more likely") for those who had most often previously used Supporting Information or the manuscript to share data. However, as this question asked about future behaviours, it should be interpreted cautiously due to social desirability bias.

Our findings show only limited support for previous studies that found that researchers are incentivised to share data by open data/science badges that act as a kind of credit/reward incentive (Kidwell et al. 2016, Rowhani-Farid et al. 2017, Pearce & Grant 2020). Researchers have said they want to receive more credit for sharing data (Digital Science et al. 2018) but our results suggest that a badge or icon is not sufficient reward to bring about wholesale behaviour change on its own. However, crucially in our case, the Accessible Data icon was added to any eligible article automatically without the authors having to opt-in or take any action they had not already taken. The Accessible Data icon is more of a tool that enhances existing data sharing behaviours by rewarding the use of repositories that provide persistent identifiers and linking of the dataset in article Data Availability Statements. However, we found when implementing the icon that many DAS link to data in repositories using a non-persistent identifier or weblink. The use of private links—which still resolve to a dataset but are less permanent than those with DOIs and intended for sharing data for peer review—in published articles is very common. There is clearly work to do to improve consistency and practice of how data links are shared, but we decided to be inclusive in how we deploy the Accessible Data icon. It displays as long readers can access the data via a link (persistent or otherwise). We decided it was more important to help researchers as authors—who may be unaware of the nuances of DOIs and private links—and also help them as readers, by including imperfect but functional links to data in our articles. This challenge of capturing the "right" link in articles that link to research data in repositories is, anecdotally, common among scholarly publishers.

Our interviews highlighted that when viewing an article as a reader, researchers generally feel confident that they would be able to find the associated data as it is usually linked from a consistent place. This suggests that an Accessible Data icon is not filling an underserved purpose, however, it is not detracting from the purpose of engaging readers with shared data. Indeed, the usage data associated with the Accessible Data icon suggests that readers did frequently engage with it. In this regard, the experiment has been effective and the reasons for this given in the interviews relate to the ease of finding links and user design rather than offering a novel solution. However, we should also remain cautious to ensure that the Accessible Data icon is not mistaken for a signal of quality for the data, as highlighted in the interviews, but merely indicates the presence of data. Additionally, the Accessible Data icon is, so far, only present on the html version of the articles and not the pdfs or in databases such as PubMed Central that index open access articles, which limits its potential visibility.

Concluding remarks and potential implications

The research findings presented here have several implications for publishers of scholarly content. Many publishers have sought to normalise or improve data sharing through policies in recent years, and promote the use of the data repositories as the preferred method for data sharing, over making data "available on [reasonable] request". General data repository integrations via Editorial Manager are available, in principle, to many journals that use this submission system but, to our knowledge, this is the first time usage of this solution has been quantified and reported formally. Our experiments show that offering a general data repository integration in Editorial Manager alone—that is, without policy or other incentives for authors—is unlikely to significantly increase the use of data repositories. Although this is a "negative" result, it is important to share to support further innovation and experimentation by publishers, repositories and infrastructure providers to improve data sharing practice. PLOS is continuing to work with Dryad and others to promote best practices in open research and this research has highlighted the importance of continued advocacy and education of authors, many of whom may have never used a repository before. Indeed, on observing low usage of Dryad in the first year, PLOS has focused more attention on promoting data repositories through educational content, such as videos[1]. Another possible outcome of this research could be to recommend a policy approach to repository use, i.e. making it mandatory, however, given it is still a minority practice, careful consideration of author experience and potential unintended consequences would need to take place.

This research has shown that the Accessible Data icon is a simple, automated, low cost solution that can significantly increase engagement with research data in repositories that are linked to published articles by 20%, and which might help to normalise the use of data repositories for data sharing. In response to these findings we have extended the scope of the Accessible Data icon to include more generalist repositories as well as specialist repositories to explore differences in engagement between disciplines and to promote direct access to data from articles (Hrynaszkiewicz 2023). This will help deepen our understanding of data access behaviours, and extend the experiment to other outputs, such as code, beyond the article. Other future developments for the Accessible Data icon might include metadata improvements to distribute the icon in more places, such as article search results and indices of journal content (such as PubMed and PubMed Central, Google Scholar) to increase its visibility. And, more sophisticated versions of the icon that characterise aspects of data interoperability and Reusability (the I and R of FAIR data).

---

[1] Videos produced in relation to this are:
1) Benefits of data sharing and repository use - https://vimeo.com/manage/videos/865129965/68315ca1bb;
2) What research data should you share - https://vimeo.com/manage/videos/865130965/ea41033c65;
3) How to prepare and share research data - https://vimeo.com/manage/videos/865132404/3f8415e15a;
4) Combined video - https://vimeo.com/manage/videos/865126608/7747b330fe

Data Availability Statement

Acknowledgements

Bibliography

Arza V, Fressoli M. 2017. Systematizing benefits of open science practices. *Information Services & Use* 37:463–474. DOI: 10.3233/ISU-170861.

Borycz J, Olendorf R, Specht A, Grant B, Crowston K, Tenopir C, Allard S, Rice NM, Hu R, Sandusky RJ. 2023. Perceived benefits of open data are improving but scientists still lack resources, skills, and rewards. *Humanities and Social Sciences Communications* 10:1–12. DOI: 10.1057/s41599-023-01831-7.

Cadwallader L, Haldar K, Malim MH. 2021. Increasing best practice data sharing at PLOS Pathogens. *PLOS Pathogens* 17:e1010021. DOI: 10.1371/journal.ppat.1010021.

Cadwallader, L, Novich, B. 2023. Supporting materials for: Incentivising best practice in research data sharing: Experiments to increase use of and engagement with data repositories. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.6913318

Campbell HA, Micheli-Campbell MA, Udyawer V. 2019. Early Career Researchers Embrace Data Sharing. *Trends in Ecology & Evolution* 34:95–98. DOI: 10.1016/j.tree.2018.11.010.

Chow NLY, Tateishi N, Goldhar A, Zaheer R, Redelmeier DA, Cheung AH, Schaffer A, Sinyor M. 2023. Does knowledge have a half-life? An observational study analyzing the use of older citations in medical and scientific publications. *BMJ Open* 13:e072374. DOI: 10.1136/bmjopen-2023-072374.

Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. 2020. The citation advantage of linking publications to research data. *PLOS ONE* 15:e0230416. DOI: 10.1371/journal.pone.0230416.

Digital Science, Goodey G, Hahnel M, Zhou Y, Jiang L, Chandramouliswaran I, Hafez A, Paine T, Gregurick S, Simango S, Palma Peña JM, Murray H, Cannon M, Grant R, McKellar K, Day L. 2022. *The State of Open Data 2022*. Digital Science. DOI: 10.6084/m9.figshare.21276984.v2.

Digital Science, Hahnel M, Fane B, Treadway J, Baynes G, Wilkinson R, Mons B, Schultes E, Luiz Olavo Bonino Da Silva Santos, Arefiev P, Osipov I. 2018. *The State of Open Data Report 2018*. Digital Science. DOI: 10.6084/M9.FIGSHARE.7195058.

Gomes DGE, Pottier P, Crystal-Ornelas R, Hudgins EJ, Foroughirad V, Sánchez-Reyes LL, Turba R, Martinez PA, Moreau D, Bertram MG, Smout CA, Gaynor KM. 2022. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences* 289:20221113. DOI: 10.1098/rspb.2022.1113.

Grant R, Pearce R. 2020. Do Open data badges influence author behaviour? A case study at Springer Nature. DOI: 10.31219/osf.io/6qsrt.

Hrynaszkiewicz I, Harney J, Cadwallader L. 2021a. A survey of code sharing practice and policy in computational biology. DOI: 10.31219/osf.io/f73a6.

Hrynaszkiewicz I, Harney J, Cadwallader L. 2021b. A Survey of Researchers' Needs and

Priorities for Data Sharing. *Data Science Journal* 20:31. DOI: [10.5334/dsj-2021-031](10.5334/dsj-2021-031).

Hrynaszkiewicz I, Simons N, Hussain A, Grant R, Goudie S. 2020. Developing a Research

Data Policy Framework for All Journals and Publishers. *Data Science Journal* 19:5.

DOI: [10.5334/dsj-2020-005](10.5334/dsj-2020-005).

Hrynaszkiewicz I. 2023.Extending Accessible Data to more articles, repositories, and

outputs. *Available at https://theplosblog.plos.org/2023/07/accessible-data/* (accessed

July 14, 2023).

Hufton A. 2015.Repository Highlight: figshare and the crucial service of generalist

repositories. *Available at*

[https://blogs.nature.com/scientificdata/2015/03/08/repository-highlight-figshare-and-the](https://blogs.nature.com/scientificdata/2015/03/08/repository-highlight-figshare-and-the-crucial-service-of-generalist-repositories/)

[-crucial-service-of-generalist-repositories/](https://blogs.nature.com/scientificdata/2015/03/08/repository-highlight-figshare-and-the-crucial-service-of-generalist-repositories/) (accessed October 23, 2023).

Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg L-S,

Kennett C, Slowik A, Sonnleitner C, Hess-Holden C, Errington TM, Fiedler S, Nosek

BA. 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective

Method for Increasing Transparency. *PLOS Biology* 14:e1002456. DOI:

10.1371/journal.pbio.1002456.

Kim Y. A study of the roles of metadata standard and data repository in science,

technology, engineering and mathematics researchers' data reuse. Online Information

Review. 2021;45: 1306–1321. doi:[10.1108/OIR-09-2020-0431](10.1108/OIR-09-2020-0431)

Krause G, Hare M, Smit M, Mongeon P. 2023. Who Re-Uses Data? A Bibliometric Analysis

of Dataset Citations. DOI: 10.48550/arXiv.2308.04379.

Lowenberg D, Chodacki J, Fenner M, Kemp J, Jones MB. 2019. *Open Data Metrics:*

Lighting the Fire. Zenodo. DOI: 10.5281/zenodo.3525349.

McIntosh LD, Sumner J, Vitale C. 2021. *Transparently Reported Research: An analysis of Wellcome-funding publications in 2016 and 2019*. Wellcome Trust. DOI: 10.6084/m9.figshare.13810220.v1.

McIntosh LD, Sumner J, Vitale C. 2022. *The State of Trust & Integrity in Research*. Ripeta. DOI: 10.6084/m9.figshare.20497062.v2.

Ni R, Waltman L. 2023. To Preprint or Not to Preprint: A Global Researcher Survey. DOI: 10.31235/osf.io/k7reb.

Novich B, Cadwallader L, Harney J. 2023. Data from Dryad Integration and Accessible Data surveys. DOI: 10.6084/m9.figshare.22762469.v1.

Pearce R, Grant R. 2020. Do Open Data Badges Influence Author Behaviour? a Case Study at Springer Nature. *International Journal of Digital Curation* 15:8. DOI: 10.2218/ijdc.v15i1.694.

Pisani E, Aaby P, Breugelmans JG, Carr D, Groves T, Helinski M, Kamuya D, Kern S, Littler K, Marsh V, Mboup S, Merson L, Sankoh O, Serafini M, Schneider M, Schoenenberger V, Guerin PJ. 2016. Beyond open data: realising the health benefits of sharing data. *BMJ* 355:i5295. DOI: 10.1136/bmj.i5295.

Piwowar HA, Vision TJ. 2013. Data reuse and the open data citation advantage. *PeerJ* 1:e175. DOI: 10.7717/peerj.175.

Public Library of Science. 2022. PLOS Open Science Indicators. DOI: 10.6084/m9.figshare.21687686.v3.

Rowhani-Farid A, Allen M, Barnett AG. 2017. What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review* 2:4. DOI: 10.1186/s41073-017-0028-9.

Serghiou S, Contopoulos-Ioannidis DG, Boyack KW, Riedel N, Wallach JD, Ioannidis JPA. 2021. Assessment of transparency indicators across the biomedical literature: How open is open? *PLOS Biology* 19:e3001107. DOI: 10.1371/journal.pbio.3001107.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M. 2011. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6:e21101. DOI: 10.1371/journal.pone.0021101.

Tenopir C, Rice NM, Allard S, Baird L, Borycz J, Christian L, Grant B, Olendorf R, Sandusky RJ. 2020. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE* 15:e0229003. DOI: 10.1371/journal.pone.0229003.

Toribio-Flórez D, Anneser L, deOliveira-Lopes FN, Pallandt M, Tunn I, Windel H. 2021. Where Do Early Career Researchers Stand on Open Science Practices? A Survey Within the Max Planck Society. *Frontiers in Research Metrics and Analytics* 5. DOI: 10.3389/frma.2020.586992.

Van den Eynden V, Knight G, Vlad A, Radler B, Tenopir C, Leon D, Manista F, Whitworth J, Corti L. 2016. Survey of Wellcome researchers and their attitudes to open research. :1843500 Bytes. DOI: 10.6084/M9.FIGSHARE.4055448.V1.

Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018. DOI: 10.1038/sdata.2016.18.