An overview of open science in eco-evo research and the publisher effect.

Khalid Belkhir, Carole M. Smadja, Pierre-Olivier Antoine, Céline Scornavacca, Nicolas Galtier*

ISEM, Univ Montpellier, CNRS, IRD, France

* Corresponding author: nicolas.galtier@umontpellier.fr

Abstract

Sharing data and code is increasingly recognized as essential for transparency and reproducibility in scientific research. This is especially critical in ecology and evolutionary biology (eco-evo), where sampling and experiments often occur in the field and can hardly be reproduced. To monitor the adoption of open science practices in eco-evo and identify key influencing factors, we analyzed 550 articles published in 2024 in 110 journals randomly sampled from the DAFNEE database. We assessed the rate of data and code sharing and examined its relationship with journal subfield, citation index, business model, partnership, and publisher. We report that about one third of eco-evo articles published in 2024 share data and/or code - excluding omics data, for which sharing is much more frequent. Data and code sharing is more prevalent in fundamental than applied fields, and shows only a weak correlation with journal citation indices. Notably, we uncovered a strong publisher effect, and report that publishers who partner with academic institutions are the ones who tend to promote open science. Our findings highlight the importance of editorial frameworks in promoting data and code sharing and offer insights into how open science can be further fostered within the eco-evo research community.

Key Words

open data; open code; reproducibility; ethical publishing

Introduction

Open science is broadly recognized to be an element of solution to the reproducibility crisis affecting science. Increasing incentive for scientists to publish in open access, share data and share code, offer other scientists the opportunity to replicate, verify, and perhaps improve their analyses (Nosek et al. 2015, Wilkinson et al. 2016). This global movement is spreading into the field of ecology and evolutionary biology (eco-evo; Hampton et al. 2015, Shaw et al. 2016, Renaut et al. 2018, Allen et al. 2019, Poisot et al. 2019, Powers et al. 2019), as illustrated by the recent launch of a learned society dedicated to this topic (https://www.sortee.org/). This is a noteworthy cultural change. If today most of us would agree that transparency is simply good scientific practice, only 10 to 15 years ago it was rare for the data or code associated with a research study to be made available alongside with the publication. There is a need to track the growth of open science in eco-evo, and to identify the factors that promote or hinder its development (Michener 2015). This is especially important in a scientific domain where data collection often occurs in the field, and for this reason is difficult to reproduce.

Although there is ample literature on the empirical assessment of open science practice in psychology (e.g. Open Science Collaboration 2015, Hardwicke et al. 2018) and medical sciences (e.g. Iqbal et al. 2016, Hamilton et al. 2023), only a limited number of such studies have concerned eco-evo. Vines et al. (2014) assessed the propensity of scientists to share data when asked by e-mail, and found that the successful response rate drops quickly with article age. Roche et al. (2015) analyzed the reusability of 100 publication-associated datasets from seven selected journals, and found that only 27% of these met the FAIR (Findable, Accessible, Interoperable, Reusable) criteria. A similar analysis conducted seven years later revealed some improvement (Roche et al. 2022, 46% reusability), while demonstrating a significant effect of the age of the principal investigator (early career scientists share more). Analyzing 229 articles from 12 selected journals, Vines et al. (2013) reported a strong influence of journal data sharing policy on authors' actual sharing practice. In this survey, 78% of the surveyed articles complied to data sharing when this was mandatory, whereas the figure was only 7% when data sharing was optional. Focusing on code sharing, Culina et al. (2020) reached a different conclusion: they found that the code sharing rate in 346 articles from 14 ecological journals was still limited (27%) despite a substantial increase in the percentage of journals with mandatory policies, compared to a previous study (Mislan et al. 2016).

The studies reviewed above provide valuable insights into the evolution of data and code sharing policy and practice over the last decade. Each of these studies, however, focused on a restricted, arbitrary subset of scientific journals. As a result, the generality of the reported results remains uncertain. Moreover, the set of target journals varies across studies, making direct comparisons perhaps hazardous. If a scientist samples an article from an eco-evo journal, what is the probability that they can access the supporting data and code? There is currently no general answer to this simple question.

Advocates of open science in eco-evo have considered various factors that might limit the rate of open science expansion, and solutions to overcome them. A major line of argument is aimed at convincing authors that sharing is beneficial, at individual or community level (Renaut et al. 2018, Allen et al. 2019, Poisot et al. 2019, Powers et al. 2019). We agree that encouraging spontaneous data and code sharing is certainly a healthy goal to pursue. Another important angle is the role played by scientific journals. Journals have the power to make open data and open code mandatory, by adopting the appropriate rules and ensuring they are followed. How efficiently eco-evo journals are moving in this direction is currently unclear (Vines et al. 2013, Culina et al. 2020). Are all journals equally committed to open science, and if not, what determines their level of commitment? Answering these questions could help researchers evaluate or prioritize journals based on the services they offer to the academic community, besides citation rates.

Here we aim to investigate data and code sharing practices across the eco-evo literature, using a random and large sample of journals and articles. By analyzing 550 articles published in 2024 in 110 eco-evo journals, we estimate the rates of data and code sharing and examine their relationship with factors such as journal subfield, business model, partnership, and publisher. Our study provides an overview of the current state of open science in eco-evo, which will serve as a milestone in monitoring this move. Importantly, we uncover a strong publisher effect, with publishers who partner with academic institutions emerging as key promoters of open science.

Material and Methods

DAFNEE

The Database of Academia-Friendly JourNals in Ecology and Evolution (DAFNEE, https://dafnee.isem-evolution.fr/) records information on the business model of scientific journals in ecology, evolutionary biology and archaeology. DAFNEE highlights >600 journals, which are non-profit journals and journals (co-)run by an academic institution such as learned societies, universities, museums and governmental research institutes. The underlying rationale is that when an article is published in a DAFNEE journal, a part of the displaced money is reinvested in academia (Racimo et al. 2022). DAFNEE encourages scientists to prioritize such journals as authors, reviewers, and editors (Galtier et al. 2025).

The main criterion of inclusion of a journal in DAFNEE is the existence of a visible partnership with an academic institution or a non-profit organization. Journals meeting this criterion are called academia-friendly. DAFNEE records a number of journal-specific features, including thematic field, publisher, academic partners, business model (subscription, hybrid, Open Access [OA], diamond OA), article processing charges (APCs), and two citations indices: the 2-year Impact Factor (IF) and the Scimago Journal Ranking citation index (SJR), which accounts for low-range citation networks (Guerrero-Bote and Moya-Anegon 2012). Note that IF is no longer visible on the DAFNEE website.

The October 2024 version of DAFNEE was here used to sample journals, stratifying by thematic field. This version includes 690 academia-friendly journals, of which 616 meet all the criteria of inclusion (Galtier et al. 2025) and are fully annotated. DAFNEE also provides a list of 488 non-academia-friendly eco-evo journals.

Journal sampling

We sampled 110 journals relevant to ecology/evolutionary biology *sensu lato*: 55 among journals that are academia-friendly (DAFNEE journals), and another 55 among journals that are entirely owned by a private publisher without any academic partnership (non-DAFNEE journals). We only

5

sampled journals publishing primary research articles in English, having an Impact Factor (IF) and an SJR index.

Academia-friendly journals in DAFNEE are assigned to 12 thematic fields. We disregarded the "General" field, which includes generalist journals publishing only a minority of eco-evo articles. The other 11 fields are "Basic ecology", "Applied ecology", "Evolution", "Zoology", "Botany", "Microbiology", "Genetics", "Health", "Palaeo/Geology", "Archaeo/Anthropology", and "Theory". We randomly sampled five DAFNEE journals per field, thus obtaining our set of 55 DAFNEE journals. The DAFNEE database also includes a list of 488 journals for which curators have not identified any academic partnership, here called non-DAFNEE journals (https://dafnee.isem-evolution.fr/, "surveyed journals" tab). Because curators do not record any thematic field for these journals, we searched the list of non-DAFNEE journals with field-specific keywords and randomly sampled five matching journals per field, thus obtaining our set of 55 non-DAFNEE journals. Three of the sampled journals (SIAM Journal of Applied Mathematics, Biology Methods and Protocols, Sedimentary Geology) contained only a minority of eco-evo articles. They were replaced with selected journals from the same field and category, namely Proceedings of the Royal Society Interface, Ecological Modelling and Cretaceous Research, respectively. For each of the sampled journals we gathered the following information: publisher, publisher type, business model, SJR, IF, and APCs. The list of sampled journals and detailed sampling procedure are available as Supplementary Material, section 1.

In addition to these 110 journals, we checked the implementation of open science in the Peer Community Journal, a journal published by the Peer Community In (https://peercommunityin.org/). Peer Community In is a recent, non-profit, Diamond Open Access initiative that promotes open science and collaboration among researchers by offering peer review, recommendation and publication of scientific articles in open access for free. To assess the reality of open science in this novel publishing platform and compare it with traditional journals, we sampled the 59 articles published by the Peer Community Journal between January and June 2024.

Annotation of journal instructions on data/code sharing

We read the instructions to authors of the 110 sampled journals and extracted information on their data policy. We qualified journal instructions using four descriptors: *opendata, openomics,*

6

opencode, *statement.* The *opendata* descriptor qualifies journal policy regarding data sharing, excluding newly generated omics data. The *openomics* descriptor qualifies journal policy regarding the sharing of newly-generated genomic, transcriptomic and proteomic data. These are often considered separately and required to be deposited on public databases such as the NCBI, EMBL or DDBJ. The *opencode* descriptor qualifies journal policy regarding code sharing. Finally, the *statement* descriptor qualifies journal policy regarding the existence of a statement on data (and code) availability within published articles.

Each of these descriptors took one of four values, *i.e.* "required", "encouraged", "mentioned", or "ignored", depending on the strength of the instructions to authors. Sharing was annotated as "required" when words like "must", "required", "mandated", or "no exception" were used. Sharing was annotated as "encouraged" when words like "should", "encouraged", or "expected" were used. When words like "can", "might", or "possible" were used the "mentioned" annotation was chosen. Finally, "ignored" was recorded when the topic was not considered at all by a journal's instructions to authors.

All the annotations were made by a single author within a two weeks period. We compared our annotations to those of Ivimey-Cook et al. (2025), who recently analyzed data- and code-sharing policies across 275 eco-evo journals. We found that the two data sets were largely consistent, and estimated the error/ambiguity rate to be around 6%. Journal annotations and detailed procedures of annotation and quality control are available as Supplementary Material, section 2.

Journal requirement score

Journal instruction annotations were summarized as a quantitative score. The data sharing requirement score of a journal was set to 3 if *opendata* was "required", 2 if "encouraged", 1 if "mentioned" and 0 if "ignored". A code sharing requirement score was similarly defined based on the *opencode* variable. The overall journal requirement score was defined as the sum of the data sharing and code sharing requirement scores. It varies between zero (no instruction) and 6 (mandatory data and code sharing). The *openomics* variable was not used to define this score since not all journals handle omics data.

Article sampling

For each journal, five primary-research articles published during the first 6 months of 2024 were randomly sampled and downloaded. Articles without any data or code, such as review/opinion articles, were discarded and replaced. For 20 hybrid journals, only OA articles were accessible to us, as our employers were not subscribers. For seven journals, there were not enough accessible articles in 2024, so that we included articles published in 2023. For each sampled article we recorded the digital object identifier (DOI), the OA/non OA status, and the country of the first mentioned affiliation. As far as the Peer Community Journal was concerned, the 59 articles published from January to June 2024 were annotated. Details on our sampling procedure and the sampled articles are available as Supplementary Material, section 3.

Annotation of article data/code-sharing

We read the sampled articles and qualified data and code sharing using seven descriptors: *has_data*, *data_availability*, *has_new_omics_data*, *new_omics_data_availability*, *has_code*, *code_availability*, *has_statement*. Below we define these descriptors and the values they can take.

has_data [yes/no]: specifies whether data were analyzed in the considered article, with the exception of newly generated omics data; *has_data* was set to "yes" when new, non-omics data were generated, or when existing data sets were re-analyzed; *has_data* was set to "no" when the only analyzed data were newly-generated omics data, or when no data at all were analyzed.

data_availability [total/partial/none/NA]: qualifies data sharing in the considered article, not considering newly generated omics data. When new non-omics data are generated, the expectation is that raw datasets are entirely shared in a reusable way. When existing data are re-analyzed, the expectation is to either precisely source the analyzed data or share files containing the analyzed data, possibly preprocessed ("cleaned"). Files can be shared either as Supplementary Material, or via a dedicated repository. Citing a reference or linking to a generalist database without specifying the associated query was not considered to be sufficiently precise sourcing. *data_availability* was set to "total", "partial" or "none" depending on whether the above requirements were entirely, partially or not at all met. Data released in a non-easily-reusable form (e.g., large tables in pdf

8

format) were considered to partially meet the requirements. *data_availability* was set to NA (missing data) when *has_data* equaled "no".

has_new_omics_data [yes/no]: specifies whether genomic, transcriptomic, or proteomic datasets were newly generated in the considered article.

new_omics_data_availability [total/partial/none/NA]: qualifies the sharing of newly-generated omics data in the considered article. The expectation is for newly-generated omics data to be deposited in a generalist database such as the NCBI, EMBL or DDBJ, and for the corresponding identifiers (accession numbers) to appear in the published article. *new_omics_data_availability* was set to NA when *has_new_omics_data* equaled "no".

has_code [yes/no]: specifies whether code was written or used to conduct the analyses realized in the considered manuscript. *has_code* was set to "yes" whenever a programming language, such as R or python, was mentioned in the Methods section, unless only minimal analyses were conducted (e.g., a correlation between two variables in R).

code_availability [total/partial/none/NA]: qualifies the sharing of code in the considered article. Files can be shared either as Supplementary Material, or via a dedicated repository. Just referring to an existing library was not considered to be significant code sharing effort. Code released in a non-easily-reusable form (e.g. in pdf format) was considered to partially meet the requirements. *code_availability* was set to NA when *has_code* equaled "no".

has_statement [yes/no]: says whether a statement on data availability appears in the considered article, irrespective of the content of the statement.

The annotations of our main data set (550 articles from the 110 DAFNEE and non-DAFNEE journals, excluding PCJ articles) were made by a single author over a six weeks period in June-July 2024. The annotations of the 59 PCJ articles were made by four authors in April 2025. A second pass was made on our main data set in December 2024, this time with the help of an automatic text extractor script. Specifically, we used the Okapi BM 25 algorithm as implemented in the PyMuPDF python library, in search for paragraphs best matching the following key words: data, code, scripts, availability, statement, share, repository, database, supplemental, supplementary, submitted. In addition, the broken links identified during the first pass were double checked. The second pass led

9

to the modification of 18 annotations. Article annotations are available in Supplementary Material, section 4. An advanced set of scripts for searching information on the sharing of research material within a scientific article is provided at https://gitlab.mbb.cnrs.fr/mbb/dafnee.

Article openness score

Article sharing annotations were summarized as a quantitative score. The openness score of an article was defined as $S=(d+c)/(D+C)$, where:

d=1 if *data_availability* equals "total", 0.5 if *data_availability* equals "partial", 0 otherwise

c=1 if *code_availability* equals "total", 0.5 if *code_availability* equals "partial", 0 otherwise

D=1 if *has_data* equals "yes", 0 otherwise

C=1 if *has_code* equals "yes", 0 otherwise.

The article openness score varies between 0 (no sharing) and 1 (total sharing), accounting for the existence of material to be shared. It was set to NA (missing data) when both *has_data* and *has_code* equaled "no". The *new_omics_data_availability* variable was not used to define this score.

Statistical analyses

Analyses were conducted with R as detailed in section 5 of the Supplementary Material.

Results and Discussion

We sampled five 2024 (rarely 2023) articles in each of 110 eco-evo journals – 10 per DAFNEE thematic field. Sixty-two journals followed a hybrid business model, 37 were OA journals, 9 diamond OA, two subscription-based. Recall that half of the sampled journals were academia-friendly, *i.e.*, had a partnership with a learned society, university, research institute or non-profit organization. Among academia-friendly journals, 22 had a for-profit publisher, 22 a non-profit publisher, and 11 were published by a University Press, while all 55 non academia-friendly journals were published for profit. The most represented publishers were Elsevier (17 journals), Springer (17) and Wiley (15). Journal APCs varied between 0 and 10,290 euros, for a mean of 2,592 euros.

10

Journal IF varied between 0.3 and 25.7 (median=3.0). SJR varied between 0.24 and 7.2 (median=1.0).

Among the 550 sampled articles, 460 were published in OA, and 90 were behind paywalls and downloaded via the bibliographic portal of *Centre National de la Recherche Scientifique* (CNRS). This is probably a higher proportion than the average publication OA rate in eco-evo. This results from the fact that, when a sampled article was not immediately accessible online or via the CNRS portal, we did not attempt to retrieve it but sampled a new one. This was the case of 20 journals. There is therefore a slight bias in favor of OA articles in this study. USA (94 articles) and China (92 articles) were by far the most common countries of first affiliation. Article first affiliations were unevenly distributed among continents: 197 were from Europe, 149 from Asia, and 115 from North-America, while each of Oceania, South America and Africa was represented by less than 40 articles. Fifteen of the sampled articles did not produce or analyze any data or generate any code, such that the effective sample size for code/data sharing analysis was 535.

Overview of eco-evo data and code sharing in 2024

Among the 487 studies that produced or analyzed data (excluding newly-generated omics data), 32% made their dataset fully available (Fig. 1A). A similar proportion was observed for code sharing: 33% of the 300 studies that relied on new code shared the relevant scripts (Fig. 1B). An additional 17% of the articles provided partial data sharing. More than half of the sampled studies did not share any bit of data or code (Fig. 1D). In contrast, data sharing via public repositories for newly-generated omics datasets was more common: 72% of the relevant studies shared omics data sets fully, and 21% partially (Fig. 1C). This stark contrast between omics and non-omics data highlights the substantial potential for improving open science practices in non-genomic fields.

It should be noted that partial or total data/code availability was recorded as soon as the relevant material was available for download at the time of article annotation. We did not check whether the files were archived in a perennial way, in a standard format, with appropriate access procedures and metadata. In other words, we assessed the Findable and Accessible properties, but not the Interoperable or Reusable ones, among FAIR principles. The proportion of FAIR eco-evo data and code in 2024 is therefore probably much below the ~1/3 figure we are here reporting – which tells about the road ahead.

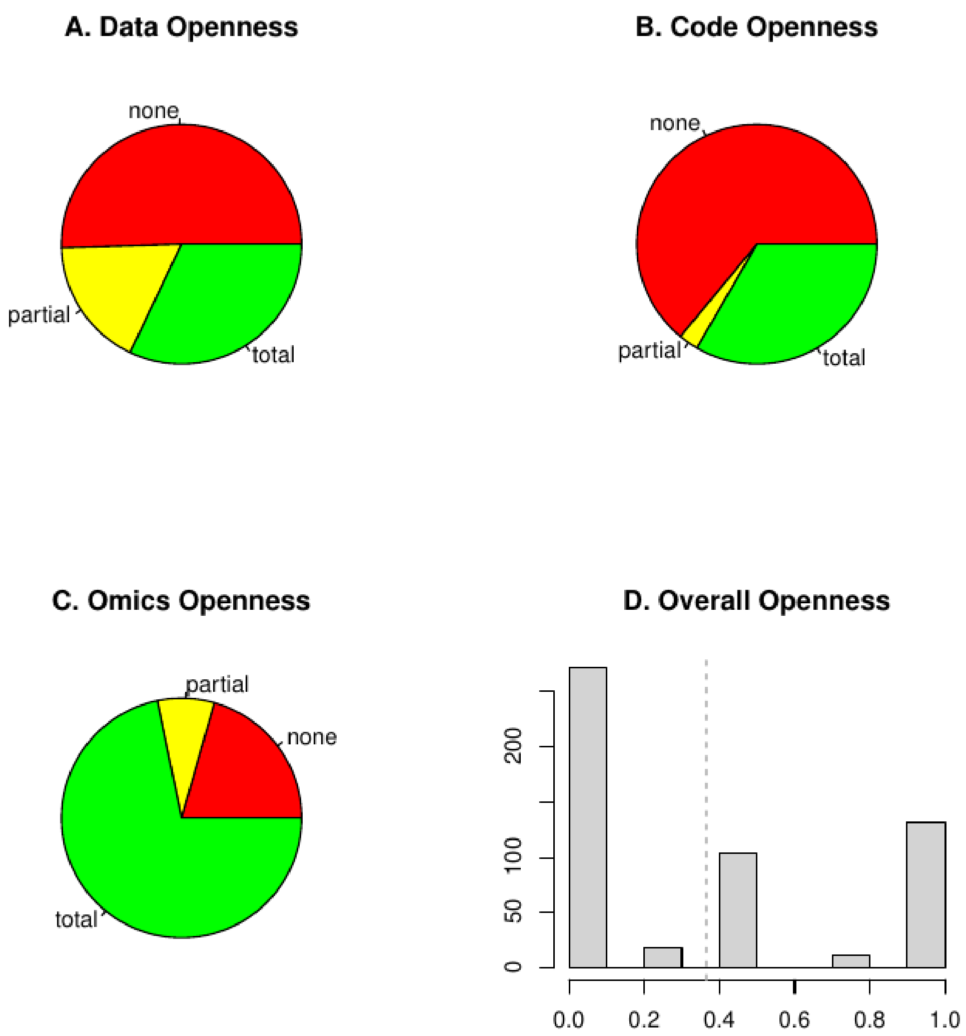

Figure1. Overview of code and data sharing prevalence in eco-evo

Distribution of non-omics data (A), code (B), omics data (C), and overall (D) openness score across 550 eco-evo articles. The overall openness score combines the data and code openness scores. It is not influenced by the omics openness score.

The reasons invoked for not sharing were varied. Among the 246 articles that produced or analyzed non-omics data but did not share files, 98 provided no justification, 88 indicated that data are available upon request to the authors, 27 argued that all the relevant data were included in the manuscript or supplementary material (but we could not find them), 10 contained a wrong or unspecific data availability statement, and eight provided a broken or useless link. Fifteen of these 246 studies indicated that data are private, mostly for ethical reasons.

12

predictor	test	*p-value*
country	Kruskal-Wallis	0.009
continent	Kruskal-Wallis	3×10^{-7}
OA	Wilcoxon	0.01
field	Kruskal-Wallis	10^{-6}
IF	Kendall	0.8
SJR	Kendall	4×10^{-5}
requirement	Kendall	5×10^{-6}
DAFNEE	Wilcoxon	2×10^{-5}
publisher	Kruskal-Wallis	6×10^{-11}

Table 1. Effect of various predictors on article openness score.

Journal-independent effects

We measured the openness of an article as the ratio of data/code sharing instances over data/code sharing opportunities, weighting partial sharing 0.5. This score disregards omics data, which are little discriminant anyway. Figure 1D shows the distribution of the openness score across articles. The score takes a discrete set of values between 0 and 1 and has a right-skewed distribution, thus departing the assumption of normality. For this reason we analyzed the determinants of its variation using non-parametric statistics.

Table 1 documents the effects of various descriptors on the article openness score. There was a significant effect of the country and continent of first affiliation (Kruskal-Wallis test, *p*-value=0.008 and 3×10^{-7}, respectively). Openness was highest for articles first-authored by scientists from North-America (mean score: 0.49), followed by Europe (0.39) and Asia (0.21). Openness also varied significantly across eco-evo fields (Kruskal-Wallis test, *p*-value=7×10^{-6}, Supplementary Fig. S1). Four fields had an average openness score below 0.3: Health, Zoology, Microbiology and Applied Ecology. Five fields had an average openness score above 0.5: Evolutionary Biology, Theory, Palaeontology, Archaeology, Basic Ecology (Supplementary Fig. S1). This suggests that data and code sharing is more common in fundamental than in applied research fields – the latter might be more closely linked to private companies. Articles published in open access tended to share data and code more often (mean score=0.39) than articles under paywalls (mean score=0.27, Wilcoxon test, p-value=0.01), implying a correlation between the various components of open science.

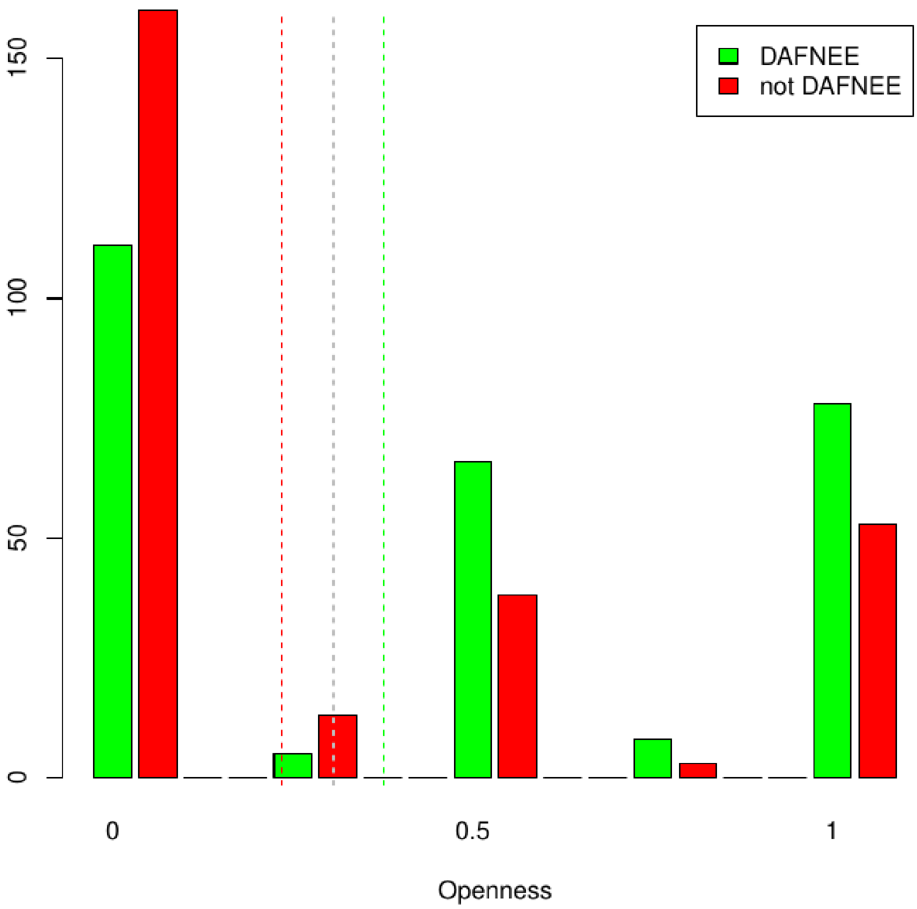

Figure 2. DAFNEE journals do more open science.

Distribution of the openness score among 225 articles published in DAFNEE-indexed journals (green) and 225 articles published in non-DAFNEE-indexed journals (red). Vertical dotted lines correspond to the distribution means: green for DAFNEE-indexed, red for non-DAFNEE-indexed, grey for all journals.

Journal-dependent effects

There was a significant effect of SJR, a citation index accounting for self-citing, on article openness score (Spearman's test, coefficient=0.18, p-value=$3x10^{-5}$). This is not particularly surprising, and might be explained by open science attracting more citations, and/or high-reputation journals being more inclined to enforce code/data sharing. We detected, however, no significant effect of journal IF on openness (Spearman's correlation coefficient=-0.01). This is, we believe, an important result, confirming previous suggestions that IF, which has been an optimization target for decades, is in no way a reliable measure of scientific quality (Fire and Guestrin 2019, Oviedo-Garcia 2021, Hanson et al. 2024). Of note, SJR in this dataset comes from a single source, the SCImago website, whereas IF data were collected from a variety of sources. This might contribute to explain the weaker relationship of the latter with article openness score.

Journal instructions to authors significantly influenced the effectiveness of code and data sharing. For example, 53% of the 169 articles published in a journal requiring data sharing did open the data fully, while this percentage was only 21% in other journals – and similarly for code sharing (53% and 27%, respectively). We summarized journal requirements regarding data sharing and code sharing as an overall requirement score. This was positively correlated with the article openness score (Spearman correlation coefficient=0.19, p-value=8×10^{-6}). These results imply that journal open science policies matter, while being insufficient in the absence of a rigorous control – nearly one half of the articles did not comply with the rules and were still published. Of note, the effect we report here is weaker than in Vines et al. (2013) – which however was based on a much smaller sample of 12 journals.

Articles published in DAFNEE journals (*i.e.*, run or co-run by an academic or non-profit organization) had a mean openness score of 0.44, which was significantly higher than the openness of articles published in entirely for-profit journals (mean score = 0.29, Wilcoxon test, p-value=2×10^{-5}). Fig. 2 shows the distribution of the openness index for DAFNEE and non-DAFNEE journals. One hypothetical explanation for this effect would be that academic institutions care about the quality and reputation of the journals they run, and therefore put pressure on editorial boards for transparency. We test this hypothesis below.

By far the strongest effect we detected was a publisher effect (Kruskal-Wallis test, p-value=6×10^{-11}). Focusing on the eight publishers for which at least three journals were sampled, we found that the mean openness score varied greatly among publishers, ranging from 0.06 (for MDPI) to 0.82 (for PLoS, Fig. 3). This was a surprising result, which indicates that publishers play a major role in the implementation of open science. Below we attempted to further characterize this effect.

15

Academia-friendly publishers promote open science

We first checked whether publishers differed in terms of the average citation rate of their journals. Indeed, the mean journal SJR varied between 0.63 and 2.4 (Fig. 3, grey dots), and the mean journal IF between 2.3 and 6.6 (Fig. 3, black diamonds) among the eight major publishers. None of these two variables, however, was significantly correlated with the publisher mean openness score. Note that publishers with a low openness score, such as MDPI, Elsevier and Frontiers Media, also have a high IF/SJR ratio, which is indicative of a tendency to self-citing. Hanson et al. (2024) provides an in-depth analysis of this phenomenon. Then we examined the variation and influence of journal open science requirements. The mean journal requirement score varied substantially among publishers, from 3 (data/code sharing usually just mentioned or encouraged) to 6 (data and code sharing always required). Indeed, we observed that the instructions to authors of many journals shared a common paragraph on open science, most likely copied from the publisher's template. The mean journal requirement score, however, was not significantly correlated with the mean openness score across publishers (Supplementary Fig. S2), which is puzzling. Frontiers media, for example, sets maximally strong rules regarding open science, but only a quarter of the articles it publishes actually share data or code. Cambridge University Press, in contrast, has generally looser requirements, but achieved 70% of data/code sharing. This suggests that publishers determine how seriously journals control the compliance with open science requirements.

We investigated the interaction between the publisher and DAFNEE effects by calculating an index of academia friendliness for each of the nine major publishers in this study. The index was defined as the ratio of the number eco-evo journals partnering with an academic or non-profit institution to the total number of eco-evo journals, for a given publisher – these two numbers being obtained by querying the DAFNEE database. The academia friendliness index varied from 0 (BMC, Frontiers Media) to 1 (PLoS). We found that the academia friendliness index was significantly and strongly correlated with the mean openness score across publishers (Fig. 4). Five publishers (BMC, Elsevier, Frontiers Media, MDPI, Springer) rarely partner with academic or non-profit institutions; they all had a mean openness score below 0.4 (mean=0.23). Four publishers (Cambridge University Press, Oxford University Press, PLoS, Wiley) co-run a majority of their journals with academic or non-profit institutions; they all had a mean openness score above 0.5 (mean=0.65). The five publishers from the first group (low academia-friendliness, low openness) are for-profit. Among the other four (high academia-friendliness, high openness), one is non-profit, two are University Presses, and one, Wiley, is for-profit.

16

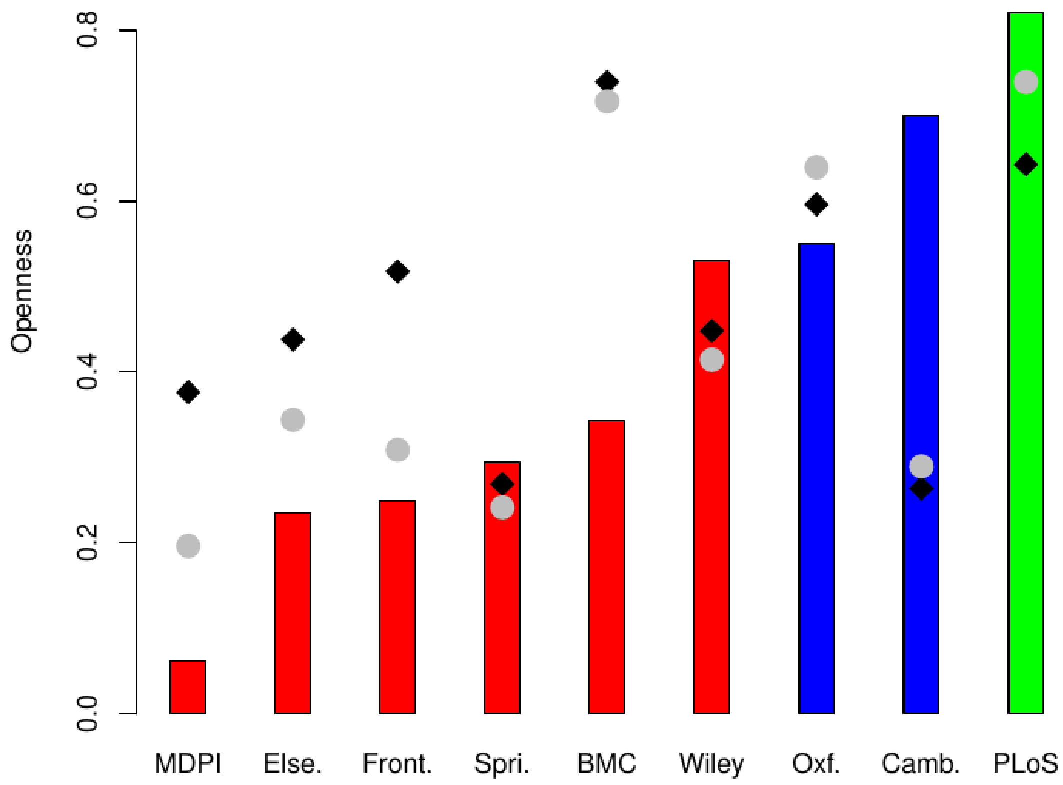

Figure 3. Openness score varies among eco-evo publishers.

Each publisher is represented by at least 15 articles. Red: for-profit publishers. Green: non-profit publisher. Blue: University Press. The coordinates of grey dots and black diamonds on the Y-axis are proportional to subfield journal average SJR and IF, respectively (scales not shown). "Else.": Elsevier; "Front.": Frontiers Media; "Spri.": Springer; "Oxf.": Oxford University Press; "Camb.": Cambridge University Press. University Presses are publishing houses associated with universities, which can be for-profit or non-profit. The two University Presses considered here are non-profit.

To better understand the drivers of this correlation, we assessed the academia-friendliness effect on openness controlling for the publisher effect. For three publishers having more than 15 sampled journals, Elsevier, Springer and Wiley, we compared the average openness score of articles published in DAFNEE vs. non-DAFNEE journals. The 15 DAFNEE journals published by Elsevier had a mean openness score (0.42) significantly higher than the 70 non-DAFNEE journals published by the same Elsevier (0.19; Wilcoxon test, *p*-value=0.02). The opposite trend, however, was found with Wiley, where the openness score was lower for DAFNEE (0.49) than non-DAFNEE (0.60) journals, although the difference was not significant. As far as Springer was concerned, DAFNEE and non-DAFNEE journals had a similar openness score of 0.29. These results only weakly support

17

the hypothesis that academic institutions deeply influence data/code sharing policy and practice. Instead, publishers appear to be the major players here: those who tend to partner with academic institutions also tend to favor open science.

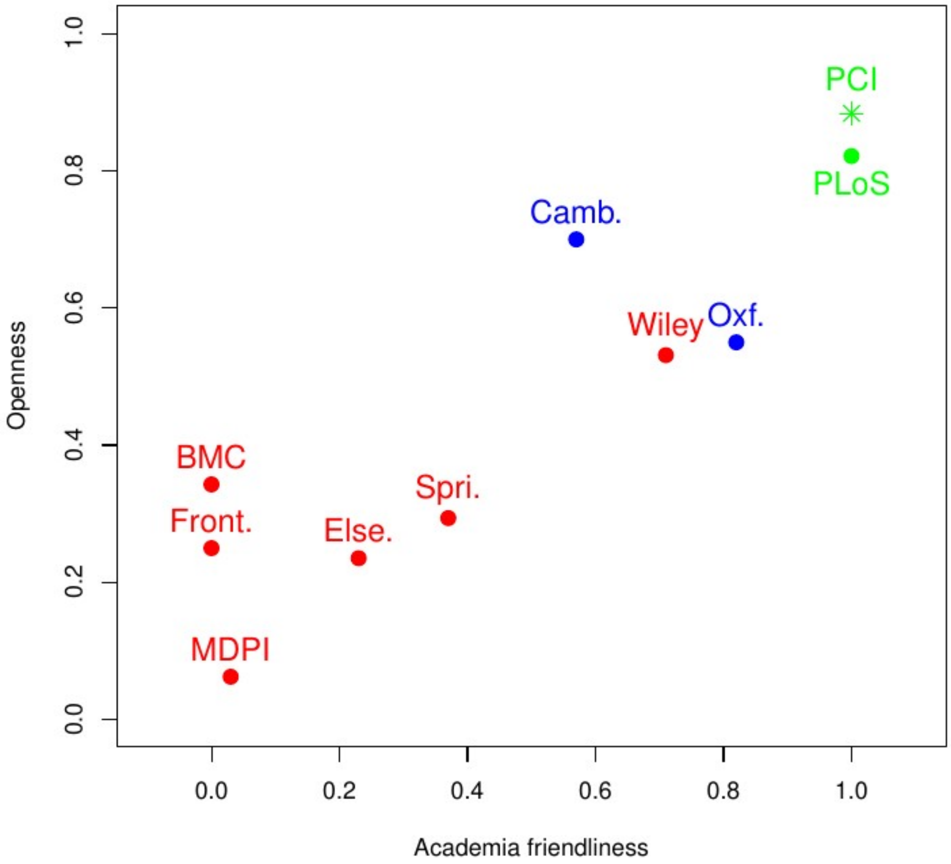

Figure 4. Academia-friendly publishers do more open science.

Y-axis: mean openness score. X-axis: percentage of published journals partnering with an academic entity. Red: for-profit. Green: non-profit. Blue: University Press. "PCI": Peer Community In. Other abbreviations: see Figure 3 caption.

Peer Community In offers a unique model where researchers can submit their work hosted on preprint repositories, to be reviewed by their peers, and eventually recommended. The authors of the recommended preprints can decide to keep their paper on the preprint server (and bypassing traditional journal structures by decoupling review and publication), send it to a traditional journal, or publish it on the Peer Community Journal. Articles published in the Peer Community Journal are open access and subject to a transparent review process, which is fully accessible on the Peer

18

Community In website. Of particular relevance to our discussion in this article is the requirement that Peer Community In managers thoroughly verify the submission of data, code, and scripts (for example, checking that the links are perennial, but also check that the information about the data/scripts/codes on the submission page and in the manuscript are consistent) of each submitted manuscript. This is done through a mandatory checklist that must be completed at submission stage. We evaluated the openness of 59 articles published by the Peer Community Journal between January and June 2024, and found that it was 0.88 on average – the highest score across all the publishers we assessed (Fig. 4).

This result further illustrates that the implementation of Open Science is a matter of will, not means – of the nine publishers shown in Fig. 4, Peer Community In has by far the smallest budget. It also highlights the importance of data editors, *i.e.*, journal editorial staff members whose role is to check the reality of data and code sharing in published articles, as described above for Peer Community In. A recent study suggests that the hiring of data editors had a strong effect on the compliance to Open Science requirements in two eco-evo journals (Ivimey-Cook et al. 2025).

Conclusions

Based on a random sample of journals and articles, we estimate that roughly one third of the eco-evo studies published in the first half of 2024 shared either data or code. This leaves substantial room for improvement and underscores the continued need to promote open science at all levels – among employers, funding agencies, scientists and journals alike. The openness score of an article was found to correlate with multiple factors, including country or continent, scientific field, SCImago Journal Ranking, Open Access status, journal requirements, academic partnership and publisher.

Two results stand out and deserve particular attention :

- The openness score is not correlated with journal impact factor. This further reinforces the message that impact factor is not a reliable indicator of scientific quality or transparency. This study adds to the growing consensus that it should be immediately abandoned as a bibliometric criterion.
- Publishers have a decisive influence on data/code sharing. The publishers with the highest openness scores are those most closely engaged with the academic community, not necessarily those with the strictest editorial rules.

Peer Community In, Cambridge University Press, Oxford University Press, PLoS and Wiley clearly do a better job than the other major publishers in promoting open data and open code. We strongly encourage eco-evo scientists to support and engage with academia-friendly journals run by these publishers - whether as authors, reviewers or editors - and especially to consider the diamond Open Access platform Peer Community In as a model of best practice.

How these results generalize outside of eco-evo is an open question. Addressing it would entail the creation of databases similar to DAFNEE in other scientific fields - an exciting prospect.

Data and code availability

Supplementary material, including the data files and source code used to generate the table and figures, are available from https://zenodo.org/records/15873598.

Author contributions

NG conceived the study, generated most of the data, conducted most of the analyses and wrote the first draft of the manuscript. KB provided analytical tools. CMS, CS and POA generated part of the data. CMS, CS, KB and POA improved the manuscript.

Funding

This work was not supported by any research grant.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgments

We thank Ed Ivimey-Cook for thoughtful comments on an earlier version of the manuscript.

References

Allen C, Mehler DMA. Open science challenges, benefits and tips in early career and beyond. PLoS Biol. 2019 May 1;17(5):e3000246. doi: 10.1371/journal.pbio.3000246. Erratum in: PLoS Biol. 2019 Dec 6;17(12):e3000587.

Culina A, van den Berg I, Evans S, Sánchez-Tójar A. Low availability of code in ecology: A call for urgent action. PLoS Biol. 2020 Jul 28;18(7):e3000763.

Fire M, Guestrin C. Over-optimization of academic publishing metrics: Observing Goodhart's Law in action. GigaScience. 2019 May 30;8(6):giz053.

Galtier N, Belkhir K, Antoine PO, Bibal C, Boëte C, Delsuc F, Huchard E, Kéfi S, Maeder C, Paradis L, Puechmaille SJ, Queffelec A, Scornavacca C, Smadja CM. Time to publish ethically: DAFNEE, a database of academia-friendly journals in ecology and evolutionary biology. EcoEvoRxiv. 2025 Jun 24;https://doi.org/10.32942/X24933

Guerrero-Bote VP, Moya-Anegón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. J Informetr. 2012 Jan 1;6(4):674–88.

Hamilton DG, Hong K, Fraser H, Rowhani-Farid A, Fidler F, Page MJ. Prevalence and predictors of data and code sharing in the medical and health sciences: systematic review with meta-analysis of individual participant data. BMJ. 2023 Jul 11;382:e075767.

Hampton SE, Anderson SS, Bagby SC, Gries C, Han X, Hart EM, Jones MB, Lenhardt WC, MacDonald A, Michener WK, Mudge J, Pourmokhtarian A, Schildhauer MP, Woo KH, Zimmerman, N. The Tao of open science for ecology. Ecosphere. 2015 Jul 23;6(7):1-13.

Hanson MA, Barreiro PG, Crosetto P, Brockington D. The strain on scientific publishing. Quant Sci Stud. 2024 Nov 1;5(4):823–43.

Hardwicke TE, Mathur MB, MacDonald K, Nilsonne G, Banks GC, Kidwell MC, Hofelich Mohr A, Clayton E, Yoon EJ, Henry Tessler M, Lenne RL, Altman S, Long B, Frank MC. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. R Soc Open Sci. 2018 Aug 15;5(8):180448.

Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible Research Practices and Transparency across the Biomedical Literature. PLoS Biol. 2016 Jan 4;14(1):e1002333.

Ivimey-Cook ER, Sánchez-Tójar A, Berberi I, Culina A, Roche DG, Almeida RA, Amin B, Bairos-Novak KR, Balti H, Bertram MG, Bliard L, Byrne I, Chan YC, Cioffi WR, CorbelQ, Elsy AD, Florko KRN, Gould E, Grainger MJ, Harshbarger AE, Hovstad KA, Martin JM, Martinig AR, Masoero G, Moodie IR, Moreau D, O'Dea RE, Paquet M, Pick JL, Rizvi T, Silva I, Szabo B, Takola E, Thoré ESJ, Verberk WCEP, Windecker SM, Winter G, Zajková Z, Zeiss R, Moran NP. From policy to practice: progress towards data- and code-sharing in ecology and evolution. EcoEvoRxiv. 2025 Jan 20;https://doi.org/10.32942/X2492Q

Michener WK. Ecological data sharing. Ecol Inform. 2015 Jul 2;29:33-44.

Mislan KAS, Heer JM, White EP. Elevating The Status of Code in Ecology. Trends Ecol Evol. 2016 Jan;31(1):4-7.

Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S, Chambers CD, Chin G, Christensen G, Contestabile M, Dafoe A, Eich E, Freese J, Glennerster R, Goroff D, Green DP, Hesse B, Humphreys M, Ishiyama J, Karlan D, Kraut A, Lupia A, Mabry P, Madon TA, Malhotra N, Mayo-Wilson E, McNutt M, Miguel E, Paluck EL, Simonsohn U, Soderberg C, Spellman BA, Turitto J, VandenBos G, Vazire S, Wagenmakers EJ, Wilson R, Yarkoni T. SCIENTIFIC STANDARDS. Promoting an open research culture. Science. 2015 Jun 26;348(6242):1422-5. doi: 10.1126/science.aab2374.

Open Science Collaboration. Estimating the reproducibility of psychological science. Science. 2015 Aug 28;349:aac4716.

Oviedo-García MÁ. Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). Res Eval. 2021 Aug 11;30(3):405–19.

Poisot T, Bruneau A, Gonzalez A, Gravel D, Peres-Neto P. Ecological data should not be so hard to find and reuse. Trends Ecol Evol. 2019 Jun;34(6):494-496.

Powers SM, Hampton SE. Open science, reproducibility, and transparency in ecology. Ecol Appl. 2019 Jan;29(1):e01822.

22

Racimo F, Galtier N, de Herde V, Bonn N, Philips B, Guillemaud T, Bourguet D. Ethical publishing: how do we get there? Philos Theory Pract Biol. 2022 Oct 21;14:15.

Renaut S, Budden AE, Gravel D, Poisot T, Peres-Neto P. Management, archiving, and sharing for biologists and the role of research institutions in the technology-oriented age. BioScience. 2018 Apr 24;68:400–1

Roche DG, Kruuk LE, Lanfear R, Binning SA. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? PLoS Biol. 2015 Nov 10;13(11):e1002295.

Roche DG, Berberi I, Dhane F, Lauzon F, Soeharjono S, Dakin R, Binning SA. Slow improvement to the archiving quality of open datasets shared by researchers in ecology and evolution. Proc Biol Sci. 2022 May 25;289(1975):20212780. doi: 10.1098/rspb.2021.2780. Epub 2022 May 18.

Shaw RG, Moore AJ, Noor M, Ritchie MG. Transparency and reproducibility in evolutionary research. J Evol Biol. 2016 Jul;29(7):1296-7.

Vines TH, Andrew RL, Bock DG, Franklin MT, Gilbert KJ, Kane NC, Moore JS, Moyers BT, Renaut S, Rennison DJ, Veen T, Yeaman S. Mandated data archiving greatly improves access to research data. FASEB J. 2013 Apr;27(4):1304-8.

Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore JS, Renaut S, Rennison DJ. The availability of research data declines rapidly with article age. Curr Biol. 2014 Jan 6;24(1):94-97.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, 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 AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, 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 SA, 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. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016 Mar 15;3:160018.

Figure captions

Figure 1. Overview of code and data sharing prevalence in eco-evo

Distribution of non-omics data (A), omics data (B), code (C), and overall (D) openness score across 550 eco-evo articles. The overall openness score combines the data and score openness scores. It is not influenced by the omics openness score.

Figure 2. DAFNEE journals do more open science.

Distribution of the openness score among 225 articles published in DAFNEE-indexed journals (green) and 225 articles published in non-DAFNEE-indexed journals (red). Vertical dotted lines correspond to the distribution means: green for DAFNEE-indexed, red for non-DAFNEE-indexed, black for all journals.

Figure 3. Openness score varies among eco-evo publishers.

Each publisher is represented by at least 15 articles. Red: for-profit publishers. Green: non-profit publisher. Red: University Press. The coordinates of black dots and black diamonds on the Y-axis are proportional to subfield journal average SJR and IF, respectively (scales not shown). "Else.": Elsevier; "Front.": Frontiers Media; "Spri.": Springer; "Oxf.": Oxford University Press; "Camb.": Cambridge University Press.

Figure 4. Academia-friendly publishers do more open science.

Y-axis: mean openness score. X-axis: percentage of published journals partnering with an academic entity. Red: for-profit. Green: non-profit. Blue: University Press. "PCI": Peer Community In. Other abbreviations: see Figure 3 caption.