ACADEMIC PRACTICE IN ECOLOGY AND EVOLUTION

Ecology and Evolution
Open Access WILEY

The data-index: An author-level metric that values impactful data and incentivizes data sharing

Amelia S. C. Hood[1] 🔵 | William J. Sutherland[1,2] 🔵

[1]Conservation Science Group, Department of Zoology, University of Cambridge, Cambridge, UK

[2]Biosecurity Research Initiative at St Catharine's (BioRISC), St Catharine's College, University of Cambridge, Cambridge, UK

Correspondence
Amelia S. C. Hood, Centre for Agri-Environmental Research, School of Agriculture, Policy and Development, University of Reading, Earley Gate, PO Box 237, Reading RG6 6AR, UK.
Email: a.s.hood@reading.ac.uk

Abstract

Author-level metrics are a widely used measure of scientific success. The h-index and its variants measure publication output (number of publications) and research impact (number of citations). They are often used to influence decisions, such as allocating funding or jobs. Here, we argue that the emphasis on publication output and impact hinders scientific progress in the fields of ecology and evolution because it disincentivizes two fundamental practices: generating impactful (and therefore often long-term) datasets and sharing data. We describe a new author-level metric, the data-index, which values both dataset output (number of datasets) and impact (number of data-index citations), so promotes generating and sharing data as a result. We discuss how it could be implemented and provide user guidelines. The data-index is designed to complement other metrics of scientific success, as scientific contributions are diverse and our value system should reflect that both for the benefit of scientific progress and to create a value system that is more equitable, diverse, and inclusive. Future work should focus on promoting other scientific contributions, such as communicating science, informing policy, mentoring other scientists, and providing open-access code and tools.

KEYWORDS

author-level metrics, bibliometrics, data citation, data metrics, data sharing, dataset repositories, FAIR research data, h-index, open science

1 | INTRODUCTION

Despite many concerns, measuring scientific success with author-level metrics has become widely used, including in funding and job allocations (Hicks & Wouters, 2015; Mingers & Leydesdorff, 2015). The h-index is the most common means of comparison; it combines the publication output (number of publications) and research impact (as number of citations) of authors (Hirsch, 2005). Despite its many flaws, the h-index has flourished because it is a simple and easily calculable measure of scientific impact. Scientists have designed similar metrics that address some of its flaws (Gasparyan et al., 2018; Mingers &

Leydesdorff, 2015), such as encouraging unwarranted self-citations (Bartneck & Kokkelmans, 2011; Senanayake et al., 2015), but these metrics are not yet as widely adopted as the h-index. Such flaws are not the focus of this paper. Here, we argue that any value system that predominantly focuses on publications will hinder scientific progress.

We need a value system that captures our wider contributions to science, both for the benefit of scientific progress and to create a value system that is more equitable, diverse, and inclusive (Davies et al., 2021). These contributions include communicating science, informing policy, mentoring other scientists, getting sponsorship, facilitating collaborations, providing open-access

code and tools, and generating useful datasets with long-term value. Some author-level metrics have been designed with this aim; Altmetrics quantifies the online impact of publications (Sud & Thelwall, 2014), and Barres (2013) suggested a metric for mentoring quality. However, many other important contributions are not yet quantified.

Here, we present a new author-level metric, the data-index, designed to value dataset output (number of datasets) and research impact (number of data-index citations) and complement other metrics of scientific success. Ignoring this dimension of scientific success results in researchers being incentivized to produce impactful publications rather than impactful datasets with long-term value. This is particularly detrimental in ecology and evolutionary biology, where datasets produced in long-term studies contribute disproportionately both scientific understanding and to policy (Hughes et al., 2017; Mills et al., 2015). Incentivizing long-term studies is especially important because they are expensive and arduous, and pressure to publish frequently is one of the key systemic barriers that hinders efforts to conduct them (Kuebbing et al., 2018). Therefore, a value system that better rewards generating impactful datasets, such as through long-term experiments, would benefit research in ecology and evolution.

Undervaluing dataset output and impact also disincentivizes data sharing. Researchers that share their data are frequently not listed as coauthors on the resulting publications. This means that the researcher's h-index is largely unaffected; if they are likely to lack what they see as sufficient credit for their contribution, they may be reluctant to share data (Ewers et al., 2019). This issue reduces opportunities for reanalysis. It is particularly prevalent in evidence synthesis, where data generators may have additional concerns that their publication will be overlooked if a synthesis paper becomes available (Patsopoulos et al., 2005; Poisot et al., 2019), and indeed, papers cited by formal review articles can experience a dramatic loss in future citations (McMahan & McFarland, 2021). In fact, synthesists have limited success in retrieving data when requested, with an estimated success rate of less than 50% if the researcher is not known personally (Côté et al., 2013; Vines et al., 2014). Synthesists contend that the citation that they give the generator's paper is sufficient credit, that generating the data does not meet authorship requirements, or that listing all generators as coauthors would be infeasible (Ewers et al., 2019). According to the Vancouver recommendations for authorship, collecting data alone does not justify authorship (ICMJE, 2019). Synthesis papers provide robust evidence for scientific theories and should be cited if relevant, and citing all of the papers within them is usually infeasible. Therefore, we clearly need a new approach. Better incentives for sharing data would also increase scientific reproducibility and save costs by avoiding the unnecessary duplication of results (Grainger et al., 2020; Piwowar et al., 2011; Wilkinson et al., 2016). The importance of data sharing is now widely recognized, and journals are increasingly mandating public data archiving (Mislan et al., 2016).

However, the majority of authors in ecology and evolution (estimated 64% by Roche et al. (2015)) archive their data in a way that prevents reuse. This demonstrates the need for an improved value system that better rewards sharing data in a useful manner.

Our call to better reward authors for producing and sharing impactful data echoes the calls of others before us, including those that developed the FAIR (Findable, Accessible, Interoperable, Reusable) Guiding Principles for data management (Ewers et al., 2019; Konkiel, 2020; Wilkinson et al., 2016). In the last decade, huge efforts have been made to increase the recognition of datasets, including the development of dataset repositories, the largest of which are DataCite and Thomson Reuters Data Citation Index (Arend et al., 2020; Cousijn et al., 2019; Konkiel, 2020; Pavlech, 2016). These repositories, respectively, store 21.8 and 10.3 million datasets as first-class research outputs ("Data Citation Index," 2021, "DataCite," 2021). Datasets are given PIDs (i.e. persistent identifiers such as DOIs), citations of these datasets are tracked, and guidelines for citing data have been developed (Cousijn et al., 2018, 2019). Citations are well-suited to showcase research impact as they are the most widely understood indicator for data, and these repositories have greatly increased the recognition of datasets (Konkiel, 2020). However, researchers are yet to be meaningfully recognized for their contributions; two recent workshops with 70 and 32 stakeholders recommended prioritizing developing metrics as a way to give credit to researchers for generating and sharing data (Federer, 2020; Pierce et al., 2019). Here, we present an author-level metric based on dataset characteristics with the aim of promoting data sharing and increasing the recognition of authors that produce impactful data. We discuss how it could be implemented by building on the existing frameworks of dataset repositories and provide user guidelines.

2 | THE DATA-INDEX

The data-index is calculated the same way as the h-index, but original datasets are ranked in order of their data-index citations rather than publications being ranked in order of publication citations (Figure 1). An author's data-index is equal to the number of datasets (n) that they have published (as primary or coauthor) that have n or more *data-index citations*. Data-index citations differ from *data citations*, which are calculated by summing the first-level citations of a dataset (Figure 2(a)) (Cousijn et al., 2018). Data-index citations are calculate by summing first- and higher-level citations (i.e. citations of datasets or publications that have cited the original dataset), but higher-level citations are only counted from datasets or publications that have reanalyzed the original dataset (Figure 2(b,c)). This means that the generators of the original dataset gain data-index citations anytime their data are cited, regardless of whether they have authored the datasets or publications that have reused their data. Though it is possible that data-index citations from the second level or higher are for datasets that are

DATA GENERATOR

(a)

	Cit.	Original data	Data used in other papers	Cit. in those papers	data-index cit.
Paper 1	18	✓	✓	21	39
Paper 2	14	✓	X	-	14
Paper 3	11	✓	✓	13	24
Paper 4	8	✓	✓	15	23
Paper 5	4	✓	X	-	4
Paper 6	4	X	-	-	-
Paper 7	1	✓	X	-	1
Paper 8	1	✓	✓	7	8
Paper 9	0	✓	X	-	0
Paper 10	0	✓	X	-	0

DATA SYNTHESIST

(b)

	Cit.	Original data	Data used in other papers	Cit. in those papers	data-index cit.
Paper 1	20	X	-	-	-
Paper 2	18	X	-	-	-
Paper 3	15	X	-	-	-
Paper 4	15	✓	✓	21	36
Paper 5	11	X	-	-	-
Paper 6	7	✓	X	-	7
Paper 7	1	X	-	-	-
Paper 8	1	X	-	-	-
Paper 9	0	✓	X	-	0
Paper 10	0	X	-	-	-

(c) *h-index = 4*

(d) *h-index = 6*
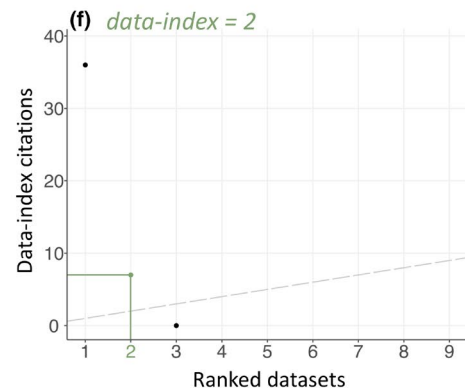
(e) *data-index = 5*
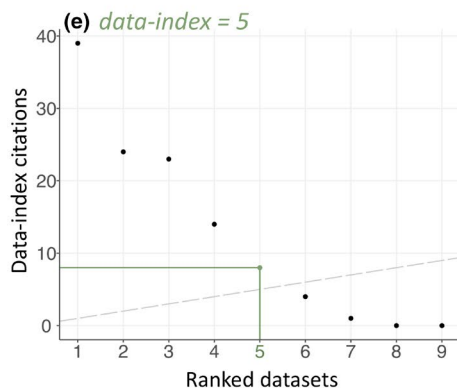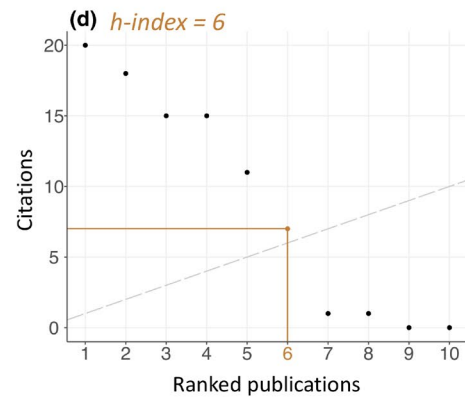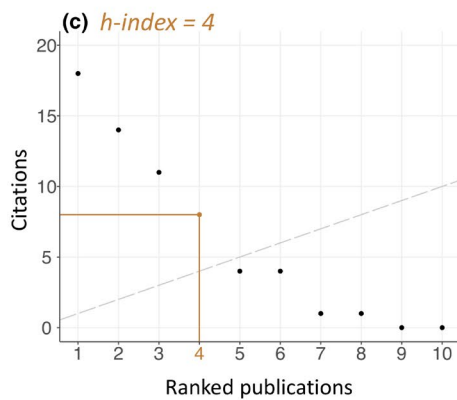
(f) *data-index = 2*

FIGURE 1 A composite figure with a hypothetical example that shows how the h-index and data-index for a data generator (i.e. someone who generates data, e.g. by conducting experiments) and data synthesist (i.e. someone who synthesizes research, e.g. through systematic reviews) at a similar career stage might differ. The h-index is equal to the number of publications (n_p) that have n_p or more citations, whereas the data-index is equal to the number of datasets (n_d) that have n_d or more data-index citations. For both indices, publications or datasets are considered the same whether the author was primary author or coauthor. (a, b) Tables showing example data used to calculate the h-index and data-index shown in plots (c–f). (a, b) Papers with original data (highlighted in gray) are the only ones included in the calculation of the data-index. Scatterplots with (c, d) publications ranked by citations to calculate the h-index and (e, f) datasets ranked by data-index citations to calculate the data-index. Dashed lines show identity lines, and colored lines show the final publication/dataset used to calculate the index value, which is also colored. In this hypothetical example, the data generator has a lower h-index (4) than the data synthesist (6), but a higher data-index (6 vs. 2). *Cit.* is an abbreviation for citation
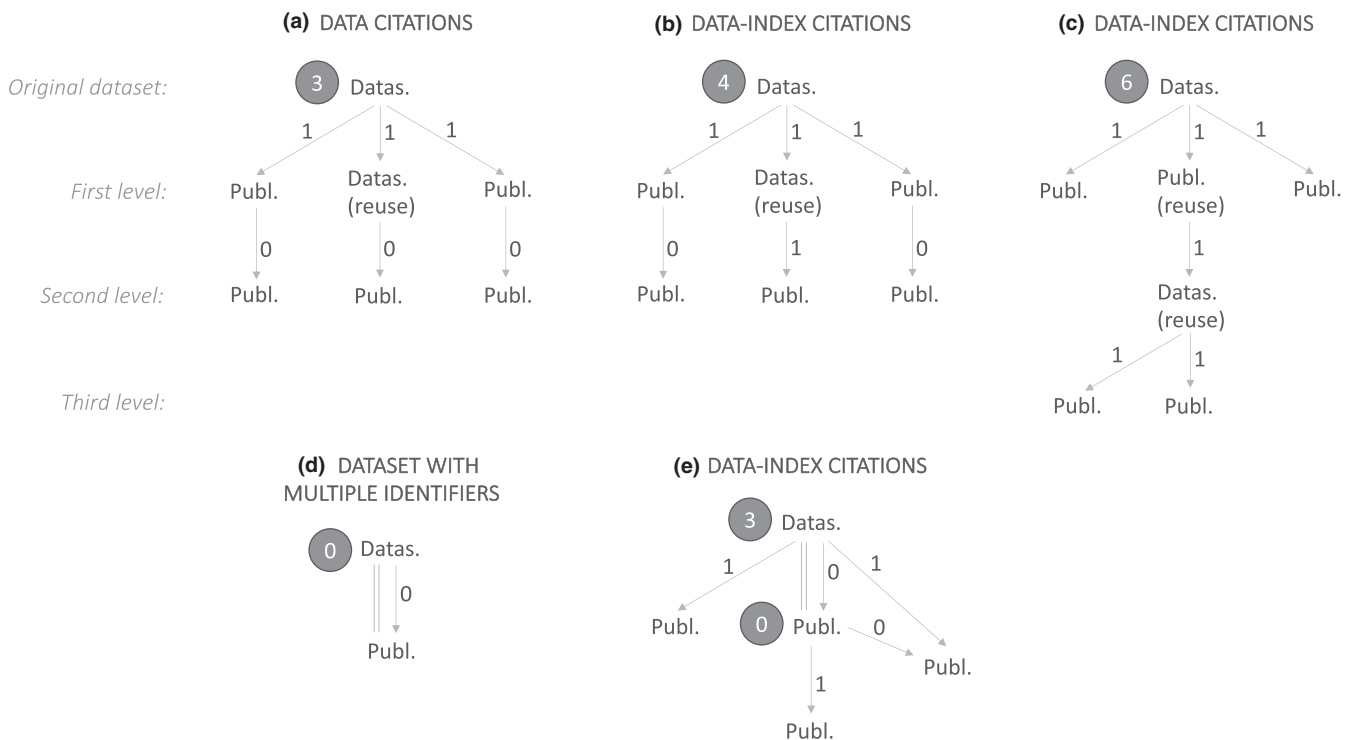
FIGURE 2 Diagrams showing that (a) data citations are calculated by summing the first-level citations of a dataset, whereas (b, c) data-index citations are calculated by summing the first-level citations of a dataset or publication that contains an original dataset and any higher-level citations of datasets or publications that have reused data from the original dataset or publication. (d, e) In cases where the same dataset has multiple identifiers (e.g. if the dataset has a unique identifier and a publication describing it has a different unique identifier), existing citation mapping software can be used to automatically group them and therefore avoid the same dataset being double-counted; parallel lines show datasets and publications that are grouped. Abbreviations are as follows: *Datas.* = dataset, *Publ.* = publication, *Datas./Publ. (reuse)* = dataset or publication that has reused data from the original dataset. Arrows show the direction of citation, and numbers in black show the value this citation gives to calculating the citation score of the original dataset. White numbers in gray circles show the (a) data citation and (b–e) data-index citation scores of the datasets beside them. Citation levels for (a–c) are shown on the left

not reusing data from the original dataset (Figure 2(c)) (for example, if a subset of data that did not include the original dataset were selected from the first-level dataset), we do not think this will be a common issue as data sharing rates are currently low. This may need to be revised if data sharing practices change. Figure 1 shows a hypothetical example of how a data generator and synthesist at a similar career stage might differ. The generator has a lower h-index than the synthesist because their publications have fewer citations, but their data-index is higher because they have more original datasets, many of which have been reused.

The data-index measures research impact (citations) rather than reuse (the number of times a dataset has been reused). For example, a dataset that is used to inform an important theory in a single highly cited paper will be considered impactful (as there are many first-level citations) even if it is not reused. However, dataset reuse (and therefore data sharing) is promoted by the data-index as the data-index citations of a dataset will increase through higher-level citations anytime the dataset is reused and cited (Figure 2). Though the impact (citations) of a paper could be unrelated to the data within it, we think that this will be in a minority of cases. Therefore, the data-index approximates how important a dataset is in informing concepts that are important within the scientific community, rather than how reusable the dataset is, but it does promote data reuse and therefore sharing.

3 | IMPLEMENTING THE DATA-INDEX

Calculating the data-index is not yet automated. Dataset repositories are developing the ability to automate calculating data-citations of the datasets in their repositories (Cousijn et al., 2019; Pavlech, 2016), and many of these datasets can be automatically linked to authors in Web of Science (Mongeon et al., 2017). We would need to build on this existing framework to include higher-level citations to automate calculating data-index citations. Another technical consideration is cases where the same original dataset is duplicated, for example, if there is a dataset with a PID and a data descriptor paper with a separate PID. Objects with different PIDs can be grouped and linked (e.g. via citation mapping software ("Connected Papers," 2021), and citation repositories are adept at doing this (Cousijn et al., 2021; Groth et al., 2020). This functionality could be used when calculating data-index citations so that the same dataset is not counted multiple times (Figure 2(d,e)).

A further consideration is datasets that do not have their own PIDs as they are only presented as part of publications. We recommend that these publications (*data studies*) are considered as original datasets because publishing datasets as first-order scientific objects is a recent practice (Robinson-García et al., 2016). Furthermore, directly citing datasets are not widely practiced (Federer, 2020) (e.g. 88% of Data Citation Index datasets have no citations (Robinson-García et al., 2016)) and even well-intentioned authors that want to cite datasets and publications in their papers, as recommended by the Joint Declaration of Data Citation Principles (Cousijn et al., 2018), may be limited by journal restrictions on reference list length. We echo the recommendation that datasets should be published as independent scientific objects to make research FAIR (Wilkinson et al., 2016), but author-level data metrics would be inaccurate if data studies were excluded because of current and historic data-publishing practices. Dataset repositories already include some data studies (e.g. there are 1.3 million data studies on the Data Citation Index ("Data Citation Index," 2021)), but wider coverage is needed to calculate complete author-level data metrics.

A final consideration is what the unit of a dataset is, and defining this is not as straightforward as defining a publication (Konkiel, 2020). It raises questions such as.

- Is there a minimum or maximum size for a dataset?
- Are two figures one dataset or two datasets?
- When does an updated dataset become a new dataset?
- Does combining datasets make a new dataset?

Answering these questions is beyond the scope of this paper as they should be addressed by groups of stakeholders, including: publishers, funders, librarians, repository administrators, open science organizations and researchers from across the disciplines. Stakeholders and data-citation experts have started this process, and developed guidelines and technical solutions to several of these issues (Cousijn et al., 2018; Federer, 2020). For example, data outputs can be grouped into single units, and the origin of datasets can be tracked in data repositories (e.g. updated datasets are labeled "isNewVersionOf") (Groth et al., 2020). We recommend a baseline of one dataset per publication, which users could deviate from according to data-citation experts' advice as it develops.

4 | GUIDELINES FOR USE

As with all author-level metrics, the data-index should be used with caution. Scientific value cannot be accurately or fairly quantified by multiple indices, but multiple indices create a more balanced view than single ones (Hicks & Wouters, 2015). We refer the reader to the Leiden Manifesto, which gives ten principles to guide research evaluation (Hicks & Wouters, 2015). We highlight four specifically: Indicators should not substitute expert assessment; research

performance should be measured against the aims of the institution or researcher; variation by field in publication and citation practices should be accounted for (Kokko & Sutherland, 1999); and false precision should be avoided.

The similarity of the h-index and the data-index has advantages and disadvantages. The data-index shares many of the limitations of the h-index (Gasparyan et al., 2018; Mingers & Leydesdorff, 2015), including using citation counts, which are biased (Davies et al., 2021). Both are liable to gaming by increasing unwarranted self-citations, and the h-index variants that correct for this could be adapted to the data-index (Bartneck & Kokkelmans, 2011; Senanayake et al., 2015). Using corrected and uncorrected metrics would give a more comprehensive view, as high rates of self-citation can be legitimate if researchers are pioneers in their field. A further concern is that both indices consider the contributions of all authors as equal, which can result in authors in successful groups having higher index scores than their counterparts in less successful groups, simply because they have coauthored many papers/datasets despite giving little input. Evaluators should consider that the data-index reinforces this bias. On balance, we think that the similarity between the h-index and data-index is beneficial, however, because its familiarity makes the limitations transparent. Citation is a concept that is well understood and valued by many in academia, and most researchers are aware of the benefits and limitations of using it for evaluations (Konkiel, 2020). Evaluators must exercise caution when using author-level metrics, which should be revised and updated as necessary (Hicks & Wouters, 2015).

5 | METRICS FOR OTHER CHARACTERISTICS OF DATASETS

The data-index calculates dataset output and research impact rather than the other characteristics of datasets, such as data usage (e.g. downloads), data reuse (e.g. times a dataset it reused), data quality (e.g. accuracy and completeness), social impact (e.g. use in informing policy), and whether the data are open access or not (Konkiel, 2020). Indicators for these characteristics are limited, and author-level metrics have not yet been developed (Konkiel, 2020). We have focussed on dataset output and impact as these characteristics reflect our current value system (publication output and impact via the h-index), and they encompass several of the most important features of data production. Some of the other characteristics will probably also be reflected in the data-index. For example, greater data usage and reuse would likely correlate with greater data impact, but data quality is not likely to be reflected in the data-index; publications with higher quality data—in terms of sample size, variance (Barto & Rillig, 2012) and replicability (Yang et al., 2020)—are not cited more than lower quality studies. Publications with open-access data are more highly cited than those without, but the mechanism for this is not yet known (Colavizza et al., 2020). Metrics that measure these characteristics directly could be quantified in other complementary indices. For example, The Google Scholar Public Access Index (Van Noorden, 2021),

which measures the proportion of an author's papers that are required by their funders to be open access are actually freely available online, could be adapted to datasets to measure data openness.

6 | DISCUSSION

The data-index addresses two major issues in the fields of ecology and evolution: that generating impactful datasets, such as long-term datasets, is undervalued, and that sharing data is disincentivized (Federer, 2020; Mills et al., 2015). Others have suggested changing authorship categories (Ewers et al., 2019) or using a standardized way to allocate coauthor contributions (e.g. via the CRediT system (Ding et al., 2021)) to better value data generators and address these issues. These approaches can be used in addition to the data-index, but they are fundamentally different as the data-index does not discriminate between the roles of coauthors in terms of their contribution to generating useful datasets. It does not specifically target data collectors, but all authors that contribute to a dataset, whether by collecting or curating the data. The data-index can be implemented retrospectively, where the other methods cannot, and it complements our existing system of using author-level metrics. The data-index answers calls from across the scientific and publishing communities to develop metrics that give credit to researchers for generating and sharing data (Federer, 2020; Pierce et al., 2019).

Our current value system (the h-index) is an oversimplification that hinders scientific progress and impact. Scientific contribution is diverse, and we need metrics that better value this diversity, both for the benefit of scientific progress and to create a value system that is more equitable, diverse, and inclusive (Davies et al., 2021). The data-index should be used to complement other author-level metrics and create a broader perspective of scientific impact. More work should be done to promote other aspects of scientific success, such as science communication (Sud & Thelwall, 2014), informing policy, mentoring scientists (Barres, 2013), getting sponsorship, facilitating collaborations, or providing open-access materials. Like many ecological and evolutionary processes, scientific success is multidimensional, and we must create a system that better values that.

CONFLICT OF INTEREST
None declared.

AUTHOR CONTRIBUTIONS
Amelia S. C. Hood: Conceptualization (lead); methodology (lead); visualization (lead); writing–original draft (lead); writing–review and editing (equal). **William J. Sutherland:** Conceptualization (supporting); methodology (supporting); writing–review and editing (equal).

DATA AVAILABILITY STATEMENT
The authors confirm that all of the data are available in the article.

ORCID
Amelia S. C. Hood https://orcid.org/0000-0003-3803-0603
William J. Sutherland https://orcid.org/0000-0002-6498-0437

REFERENCES
Arend, D., König, P., Junker, A., Scholz, U., & Lange, M. (2020). The on-premise data sharing infrastructure e!DAL: Foster FAIR data for faster data acquisition. _Gigascience_, _9_, 1–11. https://doi.org/10.1093/gigascience/giaa107

Barres, B. A. (2013). How to pick a graduate advisor. _Neuron_, _80_, 275–279. https://doi.org/10.1016/j.neuron.2013.10.005

Bartneck, C., & Kokkelmans, S. (2011). Detecting h-index manipulation through self-citation analysis. _Scientometrics_, _87_, 85–98. https://doi.org/10.1007/s11192-010-0306-5

Barto, E. K., & Rillig, M. C. (2012). Dissemination biases in ecology: Effect sizes matter more than quality. _Oikos_, _121_, 228–235. https://doi.org/10.1111/j.1600-0706.2011.19401.x

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. _PLoS One_, _15_, 1–18. https://doi.org/10.1371/journal.pone.0230416

Connected Papers [WWW Document], (2021). Retrieved from https://www.connectedpapers.com/

Côté, I. M., Curtis, P. S., Rothstein, H. R., & Stewart, G. B. (2013). Gathering data: Searching literature and selection criteria. In J. Koricheva, J. Gurevitch, K. Mengerson (Eds.), _Handbook of meta-analysis in ecology and evolution book_ (pp. 37–51). Princeton University Press. https://doi.org/10.2307/j.ctt24hq6n

Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected research: The potential of the PID graph. _Patterns_, _2_, 1–7. https://doi.org/10.1016/j.patter.2020.100180

Cousijn, H., Feeney, P., Lowenberg, D., Presani, E., & Simons, N. (2019). Bringing citations and usage metrics together to make data count. _Data Science Journal_, _18_, 1–7. https://doi.org/10.5334/dsj-2019-009

Cousijn, H., Kenall, A., Ganley, E., Harrison, M., Kernohan, D., Lemberger, T., Murphy, F., Polischuk, P., Taylor, S., Martone, M., & Clark, T. (2018). A data citation roadmap for scientific publishers. _Scientific Data_, _5_, 1–11. https://doi.org/10.1038/sdata.2018.259

Data Citation Index [WWW Document] (2021). Retrieved from https://clarivate.com/webofsciencegroup/solutions/webofscience-data-citation-index/

DataCite [WWW Document] (2021). Retrieved from https://datacite.org/

Davies, S. W., Putnam, H. M., Ainsworth, T., Baum, J. K., Bove, C. B., Crosby, S. C., Côté, I. M., Duplouy, A., Fulweiler, R. W., Griffin, A. J., Hanley, T. C., Hill, T., Humanes, A., Mangubhai, S., Metaxas, A., Parker, L. M., Rivera, H. E., Silbiger, N. J., Smith, N. S., … Bates, A. E. (2021). Promoting inclusive metrics of success and impact to dismantle a discriminatory reward system in science. _PLoS Biology_, _19_, e3001282. https://doi.org/10.1371/journal.pbio.3001282

Ding, J., Liu, C., Zheng, Q., & Cai, W. (2021). A new method of co-author credit allocation based on contributor roles taxonomy: Proof of concept and evaluation using papers published in PLOS ONE. _Scientometrics_, _126_(9), 7561–7581. https://doi.org/10.1007/s11192-021-04075-x

Ewers, R. M., Barlow, J., Banks-Leite, C., & Rahbek, C. (2019). Separate authorship categories to recognize data collectors and code developers. *Nature Ecology & Evolution*, 3, 1610. https://doi.org/10.1038/s41559-019-1033-9

Federer, L. (2020). Measuring and mapping data reuse: Findings from an interactive workshop on data citation and metrics for data reuse. *Harvard Data Science Review*, 1–18. https://doi.org/10.1162/99608f92.ccd17b00

Gasparyan, A. Y., Yessirkepov, M., Duisenova, A., Trukhachev, V. I., Kostyukova, E. I., & Kitas, G. D. (2018). Researcher and author impact metrics: Variety, value, and context. *Journal of Korean Medical Science*, 33, 1–16. https://doi.org/10.3346/jkms.2018.33.e139

Grainger, M. J., Bolam, F. C., Stewart, G. B., & Nilsen, E. B. (2020). Evidence synthesis for tackling research waste. *Nature Ecology & Evolution*, 4, 495–497. https://doi.org/10.1038/s41559-020-1141-6

Groth, P., Cousijn, H., Clark, T., & Goble, C. (2020). FAIR data reuse – The path through data citation. *Data Intelligence*, 2, 78–86. https://doi.org/10.1162/dint_a_00030

Hicks, D., & Wouters, P. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 9–11.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569–16572. https://doi.org/10.1073/pnas.0507655102

Hughes, B. B., Beas-Luna, R., Barner, A. K., Brewitt, K., Brumbaugh, D. R., Cerny-Chipman, E. B., Close, S. L., Coblentz, K. E., De Nesnera, K. L., Drobnitch, S. T., Figurski, J. D., Focht, B., Friedman, M., Freiwald, J., Heady, K. K., Heady, W. N., Hettinger, A., Johnson, A., Karr, K. A., … Carr, M. H. (2017). Long-Term studies contribute disproportionately to ecology and policy. *BioScience*, 67, 271–278. https://doi.org/10.1093/biosci/biw185

ICMJE. (2019). *Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals*.

Kokko, H., & Sutherland, W. J. (1999). What do impact factors tell us? *Trends in Ecology & Evolution*, 14, 382–384. https://doi.org/10.1016/S0169-5347(99)01711-5

Konkiel, S. (2020). Assessing the impact and quality of research data using Altmetrics and other indicators. *Scholarly Assessment Reports*, 2, 1–16. https://doi.org/10.29024/sar.13

Kuebbing, S. E., Reimer, A. P., Rosenthal, S. A., Feinberg, G., Leiserowitz, A., Lau, J. A., & Bradford, M. A. (2018). Long-term research in ecology and evolution: A survey of challenges and opportunities. *Ecological Monographs*, 88, 245–258. https://doi.org/10.1002/ecm.1289

McMahan, P., & McFarland, D. A. (2021). Creative destruction: The structural consequences of scientific curation. *American Sociological Review*, 86, 341–376. https://doi.org/10.1177/0003122421996323

Mills, J. A., Teplitsky, C., Arroyo, B., Charmantier, A., Becker, P. H., Birkhead, T. R., Bize, P., Blumstein, D. T., Bonenfant, C., Boutin, S., Bushuev, A., Cam, E., Cockburn, A., Côté, S. D., Coulson, J. C., Daunt, F., Dingemanse, N. J., Doligez, B., Drummond, H., … Zedrosser, A. (2015). Archiving primary data: Solutions for long-term studies. *Trends in Ecology & Evolution*, 30, 581–589. https://doi.org/10.1016/j.tree.2015.07.006

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, 246, 1–19. https://doi.org/10.1016/j.ejor.2015.04.002

Mislan, K. A. S., Heer, J. M., & White, E. P. (2016). Elevating the status of code in ecology. *Trends in Ecology & Evolution*, 31, 4–7. https://doi.org/10.1016/j.tree.2015.11.006

Mongeon, P., Robinson-Garcia, N., Jeng, W., & Costas, R. (2017). Incorporating data sharing to the reward system of science: Linking DataCite records to authors in the Web of Science. *Aslib Journal of Information Management*, 69, 545–556. https://doi.org/10.1108/AJIM-01-2017-0024

Patsopoulos, N. A., Analatos, A. A., & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *JAMA*, 293, 2362–2366. https://doi.org/10.1001/jama.293.19.2362

Pavlech, L. L. (2016). Data citation index. *Journal of the Medical Library Association: JMLA*, 104, 88–90. https://doi.org/10.3163/1536-5050.104.1.020

Pierce, H. H., Dev, A., Statham, E., & Bierer, B. E. (2019). Credit data generators for data reuse. *Nature*, 570, 30–32. https://doi.org/10.1038/d41586-019-01715-4

Piwowar, H. A., Vison, T. J., & Whitlock, M. C. (2011). Data archiving is a good investment. *Nature*, 473, 285. https://doi.org/10.5061/dryad.j1fd7

Poisot, T., LaBrie, R., Larson, E., Rahlin, A., & Simmons, B. I. (2019). Data-based, synthesis-driven: Setting the agenda for computational ecology. *Ideas in Ecology and Evolution*, 12, 9–21. https://doi.org/10.24908/iee.2019.12.2.e

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*, 67, 2964–2975. https://doi.org/10.1002/asi.23529

Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLoS Biology*, 13, 1–12. Https://doi.org/10.1371/journal.pbio.1002295

Senanayake, U., Piraveenan, M., & Zomaya, A. (2015). The pagerank-index: Going beyond citation counts in quantifying scientific impact of researchers. *PLoS One*, 10, 1–34. https://doi.org/10.1371/journal.pone.0134794

Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98, 1131–1143. https://doi.org/10.1007/s11192-013-1117-2

Van Noorden, R. (2021). Do you obey public-access mandates? Google Scholar is watching. *Nature*. https://doi.org/10.1038/d41586-021-00873-8

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J. S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24, 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. https://doi.org/10.1038/sdata.2016.18

Yang, Y., Youyou, W., & Uzzi, B. (2020). Estimating the deep replicability of scientific findings using human and artificial intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 10762–10768. https://doi.org/10.1073/pnas.1909046117