

## The early bird gets the return: the benefits of publishing your data sooner.

Journal:	<i>Ecology and Evolution</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Editorials
Date Submitted by the Author:	n/a
Complete List of Authors:	Lortie, Christopher; York University - Keele Campus
Category:	Community Ecology
Habitat:	Multiple
Organism:	Multiple
Approach:	Ecological Experiment
Abstract:	<p>Publish your data when collection is complete - prior to the pre-print or publication. Surface your data and meta-data openly in a data repository with a DOI and versioning. This is not an entirely novel proposal, but it is also not the norm in the historical or even contemporary culture of many scientific disciplines including ecology and evolution. Nonetheless, sharing sooner enables more rapid and effective discovery, and dramatic cultural and scientific changes can happen. The COVID-19 pandemic has highlighted the incredible power of sharing data, simple summary statistics, and all findings extremely rapidly as collected or in real-time. The public and scientific community at large used these resources to engage in open, novel data-driven science and evidence-informed decision making. In many other disciplines, we can adopt these lessons for both societal good and better science. Here, a simple set of direct benefits to the primary researcher are articulated and innovations to the scientific process and cultural practice of evidence deliberation are proposed.</p>

**The early bird gets the return: the benefits of publishing your data sooner.**

Christopher J. Lortie <sup>1\*</sup>.

1. Department of Biology, York University, Toronto, ON, Canada. M3J1P3. <https://orcid.org/0000-0002-4291-7023> \*Correspondence: [lortie@yorku.ca](mailto:lortie@yorku.ca)

Ecology and Evolution Editorial.

## Abstract

Publish your data when collection is complete - prior to the pre-print or publication. Surface your data and meta-data openly in a data repository with a DOI and versioning. This is not an entirely novel proposal, but it is also not the norm in the historical or even contemporary culture of many scientific disciplines including ecology and evolution. Nonetheless, sharing sooner enables more rapid and effective discovery, and dramatic cultural and scientific changes can happen. The COVID-19 pandemic has highlighted the incredible power of sharing data, simple summary statistics, and all findings extremely rapidly as collected or in real-time. The public and scientific community at large used these resources to engage in open, novel data-driven science and evidence-informed decision making. In many other disciplines, we can adopt these lessons for both societal good and better science. Here, a simple set of direct benefits to the primary researcher are articulated and innovations to the scientific process and cultural practice of evidence deliberation are proposed.

## Keywords

Efficiency, environmental evidence, evidence-informed decision making, open data, open science, publication process, meta-science, team science, and timing.

**Introduction**

Data are a fundamental component of the scientific process. Data can inform ideas, are one of many critical outputs of experimentation, and are the substrate of most traditional scientific reporting formats in journals in disciplines such as ecology and evolution. It has also been proposed that papers presenting more data and relatively more diverse data are more cited in ecology and evolution (Fox et al., 2016). Data and theory need not however compete for relevance in any scientific discipline including ecology (Betts et al., 2021; Fagerström, 1987; Shrader-Frechette & McCoy, 1990; Spaargaren et al., 2009). In many respects, computation and work with data can iteratively feedback to theory development and novel ideas including new experiments (Markowetz, 2017). Data, when aggregated or well articulated, can also directly and collectively inform best practices in fields such as restoration ecology in addition to the peer-reviewed publications and engender transdisciplinary solutions that involve connecting the dots through data between disciplines (Ladouceur & Shackelford, 2021). Data are thus a form of currency in many disciplines (Fan et al., 2012) but can also become a bottleneck if sharing is delayed (Westoby et al., 2021), have poorly described meta-data (Madin et al., 2007), or if there are limited incentives within the scientific culture (Michener, 2015a). Consequently, it is reasonable to propose that one should consider sharing data sooner versus later to remove or reduce the impact of these bottlenecks, particularly temporal ones, on the modern scientific process that now includes scientific communication, synthesis, and open publication as powerful final steps to connect with societal needs (Cooke et al., 2017). Full open-access journals such as Ecology and Evolution are an ideal partner that embody these open science principles including transparency and effective, timely sharing.

Hence, the following strategy is reasonable in many contexts. Publish your data when the phase of scientific collection/compilation is complete and before you begin writing the associated pre-print, report, or paper. Share the data and meta-data openly in a data repository with a DOI (digital online identifier to enable citation and permanence) and versioning (i.e. updated versions are tracked and listed) (Kenall et al., 2014; Michener, 2015b). This is not an entirely novel proposal, but it is also not the norm in the

historical or even contemporary culture of many scientific disciplines including ecology and evolution (Allen & Mehler, 2019). These two sub-disciplines of science are leaders in the advancement of many open science endeavors (Hampton et al., 2015), and big data in ecology in terms of scope and diversity of evidence thresholds have been crossed (Hampton et al., 2013). Nonetheless, sharing sooner enables more rapid and effective discovery, and dramatic knowledge and process changes can happen. The COVID-19 pandemic has highlighted the incredible power of sharing data, simple summary statistics, and all findings extremely rapidly as collected or even in real-time (George et al., 2020; Saitz & Schwitzer, 2020). The public and scientific community at large used these resources to engage in open, novel data-driven science and evidence-informed decision making (Agle, 2020; Devine et al., 2020), and these changes also innovated global learning (Lashley et al., 2020). In many other disciplines, we can adopt these lessons of being open and more rapid in sharing for both societal good and better science. Ecology and evolution are perfectly poised to embrace these changes in sharing not only publications that describe ‘bright spots’ in the anthropocene (Bennett et al., 2016) but data rapidly and openly archived for consumption, review, and reuse. This will enable new syntheses and opportunities to synthesize data associated with global change for many issues. Direct and proximate benefits to the primary researcher are proposed here, and this simple change in timing (when reasonable and viable based on your research, lab, career needs) will shift the culture of practice.

## Benefits

Global and societal benefits aside (and these are not trivial as described above), publishing data when the experiment, synthesis, or primary scientific inquiry process is complete benefits the individual researcher and collaborators in at least ten critical and distinct ways (Table 1). There are two direct classes of benefits including (i) efficiency and data improvements (quality, access, integrity, provenance, etc) and (ii) team science with societal good implications (Gabrys et al., 2016). The concept of efficiency assumes analyses, visualization, writing and peer-review need seamless and likely repeat access to the data and meta-data as we work. Workflows can be simple or complex, but it is reasonable to propose that we must

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

access and read (re-read) our data reproducibly to explore ideas - even as an individual researcher or team with our own datasets (Wilson et al., 2017). Data improvements is a more general concept predicated upon the assumption that data should be preserved in their most truthful form and then cleaned, tidied, and explored through computational tools such as R programming language (Grolemund & Wickham, 2016; Wickham, 2014). Ecoinformatic thinking is a component of most ecology and evolution, and it is not coupled to the size or complexity of a dataset or team (Michener & Jones, 2012). Team science and societal benefits are the positive outcomes associated with supporting transparency in all the steps of our work including ‘works-in-progress’ to highlight the relative importance of process versus singular product thinking for many disciplines including ecology and evolution. Sharing work sooner supports a new more inclusive culture. Full stop. Sooner enables deeper opportunities for connection with others at multiple steps in addition to review of papers (Glynn et al., 2017). This bigger picture thinking is similar to the concept of registered reports (Nosek & Lakens, 2014) and other tools that we can use to signal intent and share ideas even if we are not at the final stage of writing (if that is one of the goals of a project but there can be others). Both categories of benefits thus directly improve the research process for individuals, but the scale of benefits and implications vary by the specific benefit.

The first seven proposed benefits all stem from the anchor that data archived online, accessed remotely, and reasonably well formatted with meta-data catalyze positive opportunities for individuals and teams to both avoid errors and adopt better data management and open science practices (Table 1). Publishing sooner increases the likelihood that data will more presentable, more readable, and better tracked. This is a corollary of the Hawthorne Effect from the social sciences that proposes positive outcomes emerge from interventions including observation or participation effects (Mayo, 1949; McCambridge et al., 2014). Furthermore, code can be used to read data remotely, check for errors, do manipulations, and link data to other resources whilst keeping the original data intact (Kumuthini et al., 2020). This also enables reproducible team science and replication (Fanelli, 2018; Reed, 2018). Data repositories support versioning so tracked updates are also possible, and this promotes more transparent and accountable

science (including an opportunity to revisit previous instances). Data in a repository online are also your insurance policy against lost data, accidental deletions and propagation of these errors (Buneman et al., 2002), and for larger files, online can resolve storage challenges locally. Publishing data in most data repositories in ecology and evolution also ensures that meta-data are described appropriately and adhere to meta-data standards for a specific meta-data language required by many member nodes within the DataOne network for instance (a federation of data repositories) (Michener et al., 2012). Writing meta-data (i.e. documenting key/all aspects of data) can also become the stepping stone for the methods section in a paper. Error checking is dramatically improved through relatively longer review and access periods, additional checks by yourself and the team, and finally by publishing the data prior to peer review of the paper, you provide an opportunity for independent checks of the data by referees if they so choose (it happens). Stakeholders can also engage with the evidence, i.e. data, if openly archived, and this can precipitate novel insights. The final item directly associated with efficiency is that by sharing data online you enable the capacity for collaboration via a shared linked versus interactions and changes to a file communicated via email. Less email and tracking of files sent back and forth is a massive gain in time and integrity in your work. A brief caveat is worth mentioning. Many of these benefits of efficiency and data integrity (but not necessarily preservation) can be accrued through remote storage, privately until the team is prepared for full and open global access. Many repositories and other tools support private sharing that still provide the capacity for remote storage, easy access, and versioning. This is a viable consideration depending on the circumstances and with some forward thinking this approach can still set one up for much easier and more rapid final data publication.

The final three proposed benefits link more directly to positive outcomes for team science and societal good. Peer-review of papers can take time (Csiszar, 2016). Similar to pre-prints, publishing your data sooner can establish precedence and provide a citable object. There is also the capacity for establishing ideas sooner through not just pre-prints but via data or code to generate additional interest and use (estimated through citations or altmetrics) in the final paper (from the a project if that is the goal)

(Eysenbach, 2006; Fu & Hughey, 2019; Serghiou & Ioannidis, 2018; Shuai et al., 2012). More broadly, it accelerates collective discovery. More specifically, open data are more cited (Piwowar & Vision, 2013) and sooner extends the citation window. Sharing an idea in different forms such as a paper (written word), data, code, or visually through slide decks also promotes a more accessible and diverse set of opportunities for people to engage with concepts. Next, this is more speculative and cultural, sharing your data sooner sends a signal that you are committed to open science right the start. That is not to say that one should be disadvantaged if they cannot publish data before a paper, but that when possible, sharing and at least getting the data accessibility statement sorted for the submission of papers signals good will to editors and referees that the community can or will be able access the supporting evidence directly. Researchers can also cite their own data in the paper if available in advance (Zhao et al., 2018). Finally, data are a valid research object and an opportunity to acknowledge and provide attribution. These can be contributions and recognition associated with the data collection that may not necessarily be reflected in the authorship of the publication or report (i.e. paid technicians or others contributed to this component). Unfortunately, it is also not a given that data or evidence collected will always becomes a pre-print, paper, or written document describing and reporting on the evidence collected. Data published at this stage of a research endeavour thus represent a chance for immediate attribution. Most data repositories in ecology and evolution are primarily reviewed or checked for technical correctness and thus ensure that sharing is a viable opportunity for communicating findings in this form and recognizing participants in this process. Collectively, these latter three benefits highlight a moment in the scientific process to address more expeditious sharing and new ways to think about data publication as an instrument to shift the culture of science.

**Implications**

Now, could someone else find your data and use it? Sure. However, they are unlikely to write the same paper(s) one may or may not write about the work. The nuances of ecology and evolution and perhaps many disciplines that use a wide palette of experimental and synthesis tools likely limits the capacity for



1  
2  
3 direct experimental replication from a dataset without additional details of the methods. Nonetheless,  
4  
5 others can and will find a new use for the data. So, there is not zero risk to us of being scooped, but one  
6  
7 does risk a lost chance for attribution and provenance for ideas (and people) embodied in the collection of  
8  
9 those data. A critical implication is that we are often obligated by funders to publish data, and as  
10  
11 described above, lags in peer review can be a real challenge and publication may not always be  
12  
13 synchronous with the grant reporting cycle. Writing meta-data does consume time that is unavailable for  
14  
15 other activities like writing the paper, but we are likely to commit the time at some point to satisfy  
16  
17 requirements of agencies. Returns are also potentially higher in net efficiency and in recognition and  
18  
19 reuse that benefit the researcher and team in having more to report. The main implication is that  
20  
21 publishing sooner likely simplifies and makes your scientific life easier. It can support more efficient,  
22  
23 rapid, and open work. So, if can consider it for some of your data some of the time, make a change and  
24  
25 accelerate the research, recognition, and reuse process. A little bit more speed and openness will ensure  
26  
27 that the early bird gets the return.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Acknowledgements**

CJL is funded by an NSERC DG. There are no conflicts of interest.

**Authors 'contributions**

CJL and JF conceived the ideas; CJL and JF wrote of the manuscript. All authors gave final approval for publication.

**Data availability**

There are no data directly associated with this commentary.

For Review Only

**Table 1.** A list of proposed benefits to publishing data when the data collection process is complete. The benefits of publishing data prior to report or paper writing are developed, and two categories of benefit are proposed.

Table 1

item	benefit	description	cate
1	tidy data sooner	Publishing promotes attention to more presentable and tidy data.	effic imp
2	more readable data	Data in a repository can be sourced remotely and iteratively read into R with replicable QA/WC workflows coded.	effic imp
3	simplifies versioning and provenance	Data repositories support versioning. This ensures transparent and reproducible science. It also reduces confusion and promotes data integrity with cleaning and manipulations done from an established data asset.	effic imp
4	data are archived and online	Publishing in an established data repository is your insurance against lost data, accidental deletions, and for larger files supports distributed online and potentially collaborative analyses.	effic imp
5	meta-data are described more clearly and enables easier paper writing	Writing meta-data develops clear thinking and supports effective description of the data and easier writing of methods and results if used in a publication. Different meta-data languages engender structural thinking and can also highlight gaps in data.	effic imp
6	error checking	Individuals outside your team and authors can catch errors. Referees on papers or stakeholders can review data and meta-data and provide input on clarity.	effic imp
7	sharing within team facilitated	Data can be stored online and a link shared versus various versions of files shared by email.	effic imp
8	establishes precedence	Peer-review can take time. Similar to pre-prints, publish your data sooner can establish precedence and provide a citable object. More broadly, it accelerates collective discovery.	tean soci
9	increases likelihood of acceptance of a paper and supports open science	It is likely that open data facilitates more rapid and positive peer review of a paper using those data. It demonstrates a commitment to open science, and the authors can also cite their own data in their paper if data are online in advance.	tean soci
10	attribution	Data are a valid research object and can be assigned a citable DOI. Authorship, land acknowledgement, and funding can be recognized. Most agencies require data publication and sooner is better.	tean soci

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Review Only

## Literature cited

- Agley, J. (2020) Assessing changes in US public trust in science amid the COVID-19 pandemic. *Public Health*, **183**, 122-125.
- Allen, C. & Mehler, D.M.A. (2019) Open science challenges, benefits and tips in early career and beyond. *Plos Biology*, **17**, e3000246.
- Bennett, E.M., Solan, M., Biggs, R., McPhearson, T., Norström, A.V., Olsson, P., Pereira, L., Peterson, G.D., Raudsepp-Hearne, C., Biermann, F., Carpenter, S.R., Ellis, E.C., Hichert, T., Galaz, V., Lahsen, M., Milkoreit, M., Martin López, B., Nicholas, K.A., Preiser, R., Vince, G., Vervoort, J.M., & Xu, J. (2016) Bright spots: seeds of a good Anthropocene. *Frontiers in Ecology and the Environment*, **14**, 441-448.
- Betts, M.G., Hadley, A.S., Frey, D.W., Frey, S.J.K., Gannon, D., Harris, S.H., Kim, H., Kormann, U.G., Leimberger, K., Moriarty, K., Northrup, J.M., Phalan, B., Rousseau, J.S., Stokely, T.D., Valente, J.J., Wolf, C., & Zárrate-Charry, D. (2021) When are hypotheses useful in ecology and evolution? *Ecology and Evolution*, **n/a**.
- Buneman, P., Khanna, S., & Tan, W.-C. (2002) On propagation of deletions and annotations through views. In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 150–158. Association for Computing Machinery, Madison, Wisconsin.
- Cooke, S.J., Gallagher, A.J., Sopinka, N.M., Nguyen, V.M., Skubel, R.A., Hammerschlag, N., Boon, S., Young, N., & Danylchuk, A.J. (2017) Considerations for effective science communication. *FACETS*, **2**, 233.
- Csiszar, A. (2016) Peer review: Troubled from the start. *Nature*, **532**, 306-308.
- Devine, D., Gaskell, J., Jennings, W., & Stoker, G. (2020) Trust and the Coronavirus Pandemic: What are the Consequences of and for Trust? An Early Review of the Literature. *Political Studies Review*, **19**, 274-285.
- Eysenbach, G. (2006) Citation advantage of open access articles. *Plos Biology*, **4**, e157.
- Fagerström, T. (1987) On Theory, Data and Mathematics in Ecology. *Oikos*, **50**, 258-261.

Fan, W., Geerts, F., & Wijsen, J. (2012) Determining the Currency of Data. *ACM Trans. Database Syst.*, **37**, Article 25.

Fanelli, D. (2018) Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, **115**, 2628.

Fox, C.W., Paine, C.E.T., & Sauterey, B. (2016) Citations increase with manuscript length, author number, and references cited in ecology journals. *Ecology and Evolution*, **6**, 7717-7726.

Fu, D.Y. & Hughey, J.J. (2019) Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife*, **8**, e52646.

Gabrys, J., Pritchard, H., & Barratt, B. (2016) Just good enough data: Figuring data citizenships through air pollution sensing and data stories. *Big Data & Society*, **3**, 2053951716679677.

George, G., Lakhani, K.R., & Puranam, P. (2020) What has changed? The Impact of Covid Pandemic on the Technology and Innovation Management Research Agenda. *Journal of Management Studies*, **57**, 1754-1758.

Glynn, P.D., Voinov, A.A., Shapiro, C.D., & White, P.A. (2017) From data to decisions: Processing information, biases, and beliefs for improved management of natural resources and environments. *Earth's Future*, **5**, 356-378.

Grolemund, G. & Wickham, H. (2016) *R for Data Science* O'Reilly Media, Canada.

Hampton, S.E., Anderson, S.S., Bagby, S.C., Gries, C., Han, X., Hart, E.M., Jones, M.B., Lenhardt, W.C., MacDonald, A., Michener, W.K., Mudge, J., Pourmokhtarian, A., Schildhauer, M.P., Woo, K.H., & Zimmerman, N. (2015) The Tao of open science for ecology. *Ecosphere*, **6**, 1-13.

Hampton, S.E., Strasser, C.A., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., & Porter, J.H. (2013) Big data and the future of ecology. *Frontiers in Ecology & the Environment*, **11**, 156-162.

Kenall, A., Harold, S., & Foote, C. (2014) An open future for ecological and evolutionary data? *BMC ecology*, **14**, 10.

- Kumuthini, J., Chimenti, M., Nahnsen, S., Peltzer, A., Meraba, R., McFadyen, R., Wells, G., Taylor, D., Maienschein-Cline, M., Li, J.-L., Thimmapuram, J., Murthy-Karuturi, R., & Zass, L. (2020) Ten simple rules for providing effective bioinformatics research support. *PLOS Computational Biology*, **16**, e1007531.
- Ladouceur, E. & Shackelford, N. (2021) The power of data synthesis to shape the future of the restoration community and capacity. *Restoration Ecology*, **29**, e13251.
- Lashley, M.A., Acevedo, M., Cotner, S., & Lortie, C.J. (2020) How the ecology and evolution of the COVID-19 pandemic changed learning. *Ecology and Evolution*, **10**, 12412-12417.
- Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007) An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, **2**, 279-296.
- Markowetz, F. (2017) All biology is computational biology. *Plos Biology*, **15**, e2002050.
- Mayo, E. (1949) Hawthorne and the Western Electric Company. The Social Problems of an Industrial Civilisation. Routledge.
- McCambridge, J., Witton, J., & Elbourne, D.R. (2014) Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, **67**, 267-277.
- Michener, W.K. (2015a) Ecological data sharing. *Ecological Informatics*, **29**, 33-44.
- Michener, W.K. (2015b) Ten Simple Rules for Creating a Good Data Management Plan. *PLoS Comput Biol*, **11**, e1004525.
- Michener, W.K., Allard, S., Budden, A.E., Cook, R., Douglass, K., Frame, M., & Kelling, S. (2012) Participatory design of DataONE—Enabling cyberinfrastructure for the biological and environmental sciences. *Ecological Informatics*.
- Michener, W.K. & Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution*, **27**, 85-93.
- Nosek, B.A. & Lakens, D. (2014) Registered reports a method to increase the credibility of published results. *Soc Psychol Ger*, **45**.

Piwowar, H.A. & Vision, T.J. (2013) Data reuse and the open data citation advantage. *PeerJ*, **1**, e175.

Reed, W.R. (2018) A Primer on the 'Reproducibility Crisis 'and Ways to Fix It. *Australian Economic Review*, **51**, 286-300.

Saitz, R. & Schwitzer, G. (2020) Communicating Science in the Time of a Pandemic. *JAMA*, **324**, 443-444.

Serghiou, S. & Ioannidis, J.P.A. (2018) Altmetric Scores, Citations, and Publication of Studies Posted as Preprints. *JAMA*, **319**, 402-404.

Shrader-Frechette, K.S. & McCoy, E. (1990) Theory reduction and explanation in ecology. *Oikos*, **58**, 109-114.

Shuai, X., Pepe, A., & Bollen, J. (2012) How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. *arXiv*, 1202.2461v1.

Spaargaren, G., Mol, A., & Sonnenfeld, D. (2009) Ecological modernisation: assessment, critical debates and future directions. *The Ecological Modernisation Reader. Environmental reform in theory and practice.*, 501-520.

Westoby, M., Falster, D.S., & Schrader, J. (2021) Motivating data contributions via a distinct career currency. *Proceedings of the Royal Society B: Biological Sciences*, **288**, 20202830.

Wickham, H. (2014) Tidy Data. *Journal of Statistical Software*, **59**, 1-23.

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T.K. (2017) Good enough practices in scientific computing. *PLOS Computational Biology*, **13**, e1005510.

Zhao, M., Yan, E., & Li, K. (2018) Data set mentions and citations: A content analysis of full-text publications. *Journal of the Association for Information Science and Technology*, **69**, 32-46.