Open science and the future university researcher[1]

Kyle E. Niemeyer[2]

5 February 2017

[2] School of Mechanical, Industrial, & Manufacturing Engineering,
Oregon State University
kyle.niemeyer@oregonstate.edu
https://orcid.org/0000-0003-4425-7097

This white paper was written as a contribution to the "Imagining Tomorrow's University: Rethinking scholarship, education, and institutions for an open, networked era" workshop, a joint NIH/NSF-funded event held 8–9 March 2017 in Rosemont, IL. In this paper, I present an overview of what I consider open science, its importance, and how it plays a role in my research agenda. I also discuss challenges faced in pursuing research openness, and recommend changes to university leaders to address these barriers.

Introduction

Working openly should be the default mode of science—after all, how can we advance knowledge "by standing on the shoulders of giants"[3] if we cannot access or see those shoulders? This paper operates on the following definition of open science:

> Open science, or more broadly open research, describes the activity of performing scientific research in a manner that makes products and findings accessible to anyone. This includes sharing data openly (open data), publicly releasing the source code for research software (open-source software), and making the written products of research openly accessible (open access).

[3] "If I have seen further, it is by standing on the shoulders of giants." Isaac Newton (1676), although similar statements can be found as far back as the 12th century.

In this paper, I describe my views on why open science practices are important (supported by relevant studies), and discuss the potential societal impact of increased openness. I also summarize my policy towards open science in my research agenda. Finally, I call out pressing challenges to increased openness in research, and lay out four recommended changes for university leaders that could help encourage faculty to adopt better open practices in their research.

Importance of open science

Transforming research communities from traditional, closed environments to open ones is important for a number of reasons, including (but not necessarily limited to) the following six. McKiernan et al.[4] discuss these and additional benefits for researchers working openly. Tennant et al.[5] review in detail the benefits of open-access publications to academics and society.

[4] EC McKiernan et al. "How open science helps researchers succeed". *eLife* 5 (July 2016), e16800. DOI: 10.7554/eLife.16800.

[5] JP Tennant et al. "The academic, economic and societal impacts of Open Access: an evidence-based review [version 3; referees: 3 approved, 2 approved with reservations]". *F1000Research* 5 (Apr. 2016), p. 632. DOI: 10.12688/f1000research.8460.3.

Accessibility: Openness in research ensures that research products, particularly written output, remain accessible to all. This includes the research community, funders, policy makers, and the general public. Accessibility of research products is particularly important for publicly funded research—since the public paid for the research, the public should have access to, and be able to benefit from, it. (This does not prevent innovators or other parties from developing commercial intellectual property based on the findings, but ensures that the original discovery, when funded by the public, remains accessible to all.)

Reproducibility: Releasing products of research, including software and data, helps enable reproducibility. This is particularly true for computational science, where a written description of methods can never describe an approach as completely as source code. In general, access to research software used to perform a computational study, or the data from an experimental study, should enable others to reproduce the findings of the original researchers. However, open science is a necessary but not sufficient aspect of reproducibility, as it can be challenging to reproduce or replicate results even with available research software and data.[6]

[6] O Mesnard and LA Barba. "Reproducible and replicable CFD: it's harder than you think" (2016). arXiv:1605.04339 [physics.comp-ph]; LA Barba. "The hard road to reproducibility". *Science* 354.6308 (Oct. 2016), pp. 142–142. DOI: 10.1126/science.354.6308.142.

Impact: As a selfish motivation, performing research openly helps increase the impact of work. Studies have shown that open-access papers are cited more in most research fields. In engineering, open-access papers are cited around 1.5 times more often than non-open-access papers. Similarly, papers with associated open data were cited 9–50% more than those without.[7] Vandewalle[8] showed papers in the image-processing field receive up to three times the number of citations when source code is made available.

[7] HA Piwowar and TJ Vision. "Data reuse and the open data citation advantage". *PeerJ* 1.3 (2013), e175–25. DOI: 10.7717/peerj.175; McKiernan et al., "How open science helps researchers succeed".

[8] P Vandewalle. "Code Sharing Is Associated with Research Impact in Image Processing". *Comput. Sci. Eng.* 14.4 (June 2012), pp. 42–47. DOI: 10.1109/MCSE.2012.63.

Establish priority: Some researchers hesitate to embrace open science out of a fear of being "scooped," where competitors will use some findings, software tools, or data made available and then publish first. However, contrary to this belief, practicing open science can actually prevent being scooped: releasing preprints can establish priority of discoveries or techniques prior to the publication of a traditional peer-reviewed journal article.[9]

[9] JM Berg et al. "Preprints for the life sciences". *Science* 352.6288 (May 2016), pp. 899–901. DOI: 10.1126/science.aaf9133; C Strasser. "Preprints: The Bigger Picture". *The Winnower* 3 (July 2016), e146955.56313. DOI: 10.15200/winn.146955.56313.

The peer-review and editorial process of such papers can take many months or years, but journal articles are still necessary for research findings to be considered valid (and for researchers to receive credit). Publishing a preprint of an article publicly time-stamps the work, even as it undergoes peer review and possible revision.

Encourages trust: Embracing openness in scientific research can help encourage other researchers to trust published results, by giving the ability to inspect data or software. Soergel estimated that 5–100% of computational results given by software may be incorrect or inaccurate.[10] While simply releasing source code openly will not solve this problem, this is a necessary step towards verification and reproducibility.

[10] DAW Soergel. "Rampant software errors may undermine scientific results [version 2; referees: 2 approved]". *F1000Research* 3 (July 2015), p. 303. DOI: 10.12688/f1000research.5930.2.

It's nice: In addition to the above benefits, sharing products of research openly is kind to colleagues and the greater research community, as it prevents people from wasting time by unnecessarily repeating work. Like many graduate students, I began working on my dissertation research by attempting to reimplement another group's method and reproduce some of their results in a paper. However, this group did not share any source code openly. As a result, I wasted significant time guessing about minor implementation details or inputs not discussed in their papers. This could have been avoided by sharing the source code, which would have allowed me to more quickly move on to new work. Graduate students and other researchers constantly face similar challenges that could be avoided by greater openness in research.

Openness increases societal impact of research

Many published journal articles go unread, even in their topical domains. One study of citation rates found that 27% of papers published in the natural sciences and engineering go uncited[11,12] Those who do read most papers likely come from research institutions similar to those of the authors, even if the findings could be impactful beyond these confines, for example by leading to policy changes or technological solutions for humanitarian purposes. In part this is due to the challenging technical content, jargon, and niche topics—but it is also due to a lack of access to the journals where most research findings reside. (Making the content of these papers actually understandable or digestible by most is another challenge.)

[11] V Larivière, Y Gingras, and É Archambault. "The decline in the concentration of citations, 1900–2007". *Journal of the American Society for Information Science and Technology* 60.4 (2009), pp. 858–862. DOI: 10.1002/asi.21011.

[12] Of course, papers that are read may not be cited, and papers that are cited may not actually be read.

Considering the high and ever-increasing cost of scholarly journal subscriptions, research results should not be limited to those with the means to purchase access. By self-archiving (i.e., green open access) or publishing articles in open-access journals, researchers can ensure access for all members of society, including policymakers, funders, members of the media, entrepreneurs, and the general public—or scientists and engineers in the Global South.

Furthermore, being more open with all outputs of research (e.g., papers, software, data) could help improve the general public's perception and trust in scientific research. Simply making research products available will not solve all of these problems—for one, it will not sway those who strongly believe ideas contrary to fact. However, ensuring everyone has access to the data researchers generate and analyze, and the software tools on which we rely, could eliminate one major barrier to trust in our findings.[13]

[13] A Grand et al. "Open Science: A New 'Trust Technology'?" *Science Communication* 34.5 (2012), pp. 679–689. DOI: 10.1177/1075547012443021.

Open science and my research agenda

Over the past year and a half (September 2015–February 2017), as I initiated and built my research group, I worked to crystallize my vision and plan for performing research openly. The guiding principles I formed were heavily inspired by Lorena Barba's Reproducibility PI Manifesto,[14] the Peer Reviewers' Openness Initiative,[15] and others. However, in addition to these exemplars, my goals are also motivated by the desire to work more openly than the norm in my field.

[14] LA Barba. *Reproducibility PI Manifesto*. figshare. 2012. DOI: 10.6084/m9.figshare.104539.v1.

[15] RD Morey et al. "The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review". *Royal Society Open Science* 3.1 (2016). DOI: 10.1098/rsos.150547.

My research field includes computational modeling of combustion and fluid dynamics, chemical kinetics, and associated numerical methods, and lacks reputable open-access journals. As a graduate student, I began submitting preprints of my publications to the arXiv[16], and later archived my (otherwise non-accessible) conference papers on Figshare[17].

[16] https://arxiv.org

[17] https://figshare.com

Since then, I have adopted the following policy for my group's research:

- All written research products are made openly accessible, either through green or gold open-access avenues. Since my field generally lacks recognized, fully open-access journals, this objective is typically met by submitting preprints to services such as arXiv, engrXiv, PeerJ Preprints, or Figshare, depending on the topic. Conference papers, when not submitted to an open venue, are also made openly available. Where possible, all preprints will be released under the Creative Commons Attribution (CC BY) license.[18] (I choose, in general, to not follow the hybrid gold open-access model by paying a non-fully open journal to make a paper accessible.)

[18] https://creativecommons.org/licenses/by/4.0/legalcode

- All new research software developed by my group will be done so openly (e.g., on GitHub) and released publicly under a permissive license, such as the BSD 3-clause license. The Git version-control system is used to track the history of software projects, and releases associated with publications or data are archived (with DOIs) using Zenodo[19]. In addition, we attempt to describe as many implementation details as necessary to reproduce our work.

[19] https://zenodo.org

- All data generated through research, when serving as the basis for a publication, is archived publicly and cited appropriately in our work. Figures and plotting scripts are shared via the CC BY license and cited in our papers.

This policy will be implemented in part by incorporating its content into funding proposals, for example in Data Management Plans.

In addition to these standards I established for my group's research, I have also gotten involved with and helped initiate some community efforts to encourage and reward open science practices.

First, after participating in the working group on software credit at the Third Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE3),[20] I joined the FORCE11 Software Citation Working Group as a cochair, and helped coauthor the resulting Software Citation Principles.[21] One purpose of standardizing software citation is to help ensure authors/developers receive academic credit for their work in releasing open research software. I also serve on the editorial boards of the *Journal of Open Research Software*[22] and *Journal of Open Source Software*[23], which offer publication venues focused solely on open-source research software.

I have also collaborated with colleagues in engineering fields to improve venues for open-access publications. I serve on the editorial board of *The Journal of Open Engineering*, and published an editorial discussing the importance of open publishing in engineering.[24] With some of the same collaborators, I helped initiate engrXiv[25], an open archive for engineering publications, inspired by the arXiv.

[20] DS Katz et al. "Report on the Third Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE3)". *Journal of Open Research Software* 4 (Jan. 2016), e37. DOI: `10.5334/jors.118`.

[21] AM Smith et al. "Software citation principles". *PeerJ Computer Science* 2 (Sept. 2016), e86. DOI: `10.7717/peerj-cs.86`.

[22] `http://openresearchsoftware.metajnl.com/`

[23] `http://joss.theoj.org/`

[24] D Berg, KE Niemeyer, and L Fleischfresser. "[Editorial] Open Publishing in Engineering". *The Journal of Open Engineering* (2016). URL: `http://www.tjoe.org/article/open-publishing-in-engineering`.

[25] `http://engrxiv.org`

Challenges to performing open science

In my view, the challenges impeding greater adoption of open-science practices are mainly institutional and cultural, rather than technical. General venues for sharing and developing the products of research openly abound these days, with the availability of services like arXiv, engrXiv, and PeerJ Preprints for ensuring open access of publications; repositories like GitHub for developing (and version-controlling) research software openly; and data and software archives like Zenodo and Figshare, which practically have no file size limitations.[26] Of course, some technical problems remain: How do we make results of computational science, particularly when it involves demanding high-performance computing resources, truly reproducible? How can we cite software and data consistently, when the version might change regularly?

[26] Zenodo currently accepts datasets up to 50GB, but stores data in the CERN Data Center, along with 100PB of physics data from the Large Hadron Collider (`https://zenodo.org/faq`).

Instead, cultural inertia and lack of institutional recognition/rewards pose the most significant challenges to increased openness in science.

In my opinion, the biggest barrier to greater openness in research is ambivalence or outright hostility in many research communities. Many academic researchers either disagree on or are unaware of the importance (and benefits) of working openly. Since they were not trained in doing this, e.g., during graduate school or during postdoctoral training, they also may simply be unaware of how to research openly, or the resources that are available to do so. Furthermore, since most of their colleagues, collaborators, and competitors do not practice open science, no pressure comes from the research community to change. In addition, some communities do not support or actively oppose activities such as submitting preprints.

This lack of pressure is related to the other major issue: lack of institutional recognition and reward for open practices. In general, academic researchers will work on what gets them credit for promotion and tenure—anything beyond that requires strong intrinsic motivation, or external motivators from the research community. At most institutions, promotion and tenure review includes some judgement (whether explicit or implicit) of where faculty publish their work, but many, "high-impact"' traditional publication venues—particularly domain journals—may not support, e.g., the posting of preprints.

Recommendations for university leaders

The following list of recommendations are mostly targeted at changing criteria for promotion and tenure, and performance reviews, to encourage faculty to practice more open science:

- Consider accessibility/openness of research products along with quantity and "quality" in promotion and tenure review. Mandate self-archiving of publications (i.e., green open access).
- Recognize research products such as software and data, and their associated impacts (e.g., citations), as equal to traditional publications in scholarly impact.
- Reduce the importance of publishing in traditional venues for promotion and tenure, recognizing these may be barriers to openness.
- Support efforts to teach undergraduate and graduate students about open science skills, and those necessary to work with software and data, with the same enthusiasm that traditional lab courses receive.

Research communities that impede openness cannot be forced to change from the outside. Instead, by making changes to institutional award systems, researchers will be encouraged to improve their open practices, and thus evolve communities from the inside.

References

Barba, LA. *Reproducibility PI Manifesto*. figshare. 2012. DOI: 10.6084/m9.figshare.104539.v1.

– "The hard road to reproducibility". *Science* 354.6308 (Oct. 2016), pp. 142–142. DOI: 10.1126/science.354.6308.142.

Berg, D, KE Niemeyer, and L Fleischfresser. "[Editorial] Open Publishing in Engineering". *The Journal of Open Engineering* (2016). URL: http://www.tjoe.org/article/open-publishing-in-engineering.

Berg, JM, N Bhalla, PE Bourne, M Chalfie, DG Drubin, JS Fraser, CW Greider, M Hendricks, C Jones, R Kiley, S King, MW Kirschner, HM Krumholz, R Lehmann, M Leptin, B Pulverer, B Rosenzweig, JE Spiro, M Stebbins, C Strasser, S Swaminathan, P Turner, RD Vale, K VijayRaghavan, and C Wolberger. "Preprints for the life sciences". *Science* 352.6288 (May 2016), pp. 899–901. DOI: 10.1126/science.aaf9133.

Grand, A, C Wilkinson, K Bultitude, and AFT Winfield. "Open Science: A New 'Trust Technology'?" *Science Communication* 34.5 (2012), pp. 679–689. DOI: 10.1177/1075547012443021.

Katz, DS, SCT Choi, KE Niemeyer, J Hetherington, F Löffler, D Gunter, R Idaszak, SR Brandt, MA Miller, S Gessing, ND Jones, N Weber, S Marru, G Allen, B Penzenstadler, CC Venters, E Davis, L Hwang, I Todorov, A Patra, and M de Val-Borro. "Report on the Third Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE3)". *Journal of Open Research Software* 4 (Jan. 2016), e37. DOI: 10.5334/jors.118.

Larivière, V, Y Gingras, and É Archambault. "The decline in the concentration of citations, 1900–2007". *Journal of the American Society for Information Science and Technology* 60.4 (2009), pp. 858–862. DOI: 10.1002/asi.21011.

McKiernan, EC, PE Bourne, CT Brown, S Buck, A Kenall, J Lin, D McDougall, BA Nosek, K Ram, CK Soderberg, JR Spies, K Thaney, A Updegrove, KH Woo, and T Yarkoni. "How open science helps researchers succeed". *eLife* 5 (July 2016), e16800. DOI: 10.7554/eLife.16800.

Mesnard, O and LA Barba. "Reproducible and replicable CFD: it's harder than you think" (2016). arXiv:1605.04339 [physics.comp-ph].

Morey, RD, CD Chambers, PJ Etchells, CR Harris, R Hoekstra, D Lakens, S Lewandowsky, CC Morey, DP Newman, FD Schönbrodt, W Vanpaemel, EJ Wagenmakers, and RA Zwaan. "The Peer Reviewers' Openness Initiative: incentivizing open research practices through peer review". *Royal Society Open Science* 3.1 (2016). DOI: 10.1098/rsos.150547.

Piwowar, HA and TJ Vision. "Data reuse and the open data citation advantage". *PeerJ* 1.3 (2013), e175–25. DOI: `10.7717/peerj.175`.

Smith, AM, DS Katz, KE Niemeyer, and FORCE11 Software Citation Working Group. "Software citation principles". *PeerJ Computer Science* 2 (Sept. 2016), e86. DOI: `10.7717/peerj-cs.86`.

Soergel, DAW. "Rampant software errors may undermine scientific results [version 2; referees: 2 approved]". *F1000Research* 3 (July 2015), p. 303. DOI: `10.12688/f1000research.5930.2`.

Strasser, C. "Preprints: The Bigger Picture". *The Winnower* 3 (July 2016), e146955.56313. DOI: `10.15200/winn.146955.56313`.

Tennant, JP, F Waldner, DC Jacques, P Masuzzo, LB Collister, and CHJ Hartgerink. "The academic, economic and societal impacts of Open Access: an evidence-based review [version 3; referees: 3 approved, 2 approved with reservations]". *F1000Research* 5 (Apr. 2016), p. 632. DOI: `10.12688/f1000research.8460.3`.

Vandewalle, P. "Code Sharing Is Associated with Research Impact in Image Processing". *Comput. Sci. Eng.* 14.4 (June 2012), pp. 42–47. DOI: `10.1109/MCSE.2012.63`.