CrossMark

EDITORIAL

The Information Sharing Statement Grows Some Teeth

David N. Kennedy[1]

"Reproducibility, reproducibility, reproducibility", like the real estate mantra "location, location, location", is fast becoming a point of major emphasis with scientific funding agencies. Given the importance of the concept of reproducibility in the scientific method, it can a little surprising to some that such a fundamental concept has become such a problem that it needs to be mantrafied. Without delving again into how we got to this problem (it has been well documented in other places[1][2][3]), it is worth recounting one of the mantras that has come before, namely, "publish, publish, publish" (and it's corollary "publish or perish"). Historically, researchers have done a fabulous job of writing papers (and getting the attendant perks of grants and promotions). The translation of research findings into clinical benefits to society, however, particularly in some areas, has been less exceptional[4]

In the quest to make the research article more reproducible, *Neuroinformatics* has led the way in a number of initiatives. Promoting data and software publication[5] and requirement for an inclusion of an Information Sharing Statement (ISS)[6] are examples of these initiatives. Originally optional, and later becoming required, the ISS was designed to provide an explicit statement of if, how, and where the data, software, model and other resources used in a publication were available. While not requiring that these resources be available, the ISS required the availability of these resources to be documented, thereby attempting to exert some subtle pressure when resources were not available or only available 'upon request', practices that have been documented to have a rather poor track record for actual successful resource sharing. The time for subtlety, however, is over. Authors, journals, funders, reviewers, publishers, etc. must continue to take all elements of reproducibility even more seriously. Starting in 2018, for all articles appearing in *Neuroinformatics* we will require the following: data used must be reposited in a domain-appropriate repository with the minimum necessary barriers to access by others; and software used must be accessible from an appropriate distribution system and documented by version. Unique identifiers (RRID, DOI, accession numbers, etc.) as well as literature citation should be used to explicitly document all resources used in the publication.

The sufficiency of a submission in meeting these mandates is to be evaluated and adjudicated by the reviewers and editors

[1] Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. Doi:10.1038/nrn3475

[2] Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8), e124. Doi:10.1371/journal.pmed.0020124

[3] Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9), 712–712. Doi: 10.1038/nrd3439-c1

[4] Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17(12), 1174–1179. https://doi.org/10.1038/mp.2012.105

✉ David N. Kennedy
   David.Kennedy@umassmed.edu

1   University of Massachusetts Medical School, Worcester, MA, USA

[5] De Schutter, E., Ascoli, G. A., & Kennedy, D. N. (2009). Review of papers describing neuroinformatics software. *Neuroinformatics*, 7(4), 211–212. Doi: 10.1007/s12021–009-9058-x

[6] Kennedy, D. N. (2004). Barriers to the socialization of information. *Neuroinformatics*, 2(4), 367–368. Doi: 10.1385/NI:2:4:367

for compliance with respect to the available standards and resources of the specific subdomain covered by an article. With respect to data, it is not necessary for the journal to dictate specific data hosting facilities, but rather for the specific community to recognize a host as appropriate for that data type. There are a plethora of general (https://figshare.com/, http://datadryad.org/, etc.) and domain specific repositories (see[7] for neuroimaging examples). Access barriers to data reuse are sometimes necessary in order to assure human subjects ethics compliance, but the onerousness of these barriers must be kept to the minimum necessary. Requirements for citation of the data source (for appropriate credit to data acquirers and funders) is expected,[8] and the process of data archival should adhere to the FAIR standards (findable, accessible, interoperable and reusable)[9] and include provisions for unique and permanent citation.

For software used in publications, all software should be made available through some standard distribution system (i.e. https://github.com/). Even software that has not been fully prepared for sharing (i.e. with limited documentation, no user support, limited platform support, etc.) should be made available to others if it is basis for a publication. Best practices in software development (documentation, unit testing, etc.) can be promoted, but code accessibility must be required. As with data, software should have specific and unique identifiers available,[10] indicate version, and provide information about the execution environment used.

As part of the considerations of the citation of software and data, at *Neuroinformatics* we actively promote the disambiguation of 'use' of a resource in the article compared to the mentioning of a resource in the context of discussion. For example, the statement "We considered Resource A (citation A) and Resource B (citation B) but chose to use Resource C (citation C)" includes valuable citation credit for all three resources, but lack attribution of an even more valuable credit for being the resource actually used. Simple citation counting (and h-index generation), cannot specifically distinguish between referencing and using. In order to help make this important distinction, both in terms of research author credit and to increase reproducibility, we have adopted the use of Research Resource Identifiers (RRID)[11] as part of the ISS. RRIDs are a human readable unique identifier associated with the concept of a resource (a particular software program, or a particular antibody, etc.) and the citation of these identifiers is designed to indicate specific use of that resource. This helps resource providers to expand upon a citation index for a resource (a given resource may have many publications that could have been cited, and only some unknowable subset of these citations are for actual use of the resource), to include a specific 'use-index' by counting RRID citations to a particular resource. In addition, finding all publications that *use* a specific resource (in order to aggregate similar results, etc.) is simplified.

Enhancing the ability of authors to document exactly what has been done, and facilitating the reader's ability to actively engage in the reproducibility process is a core functionality of the publication process. As the technology to enhance the ability of any given publication to be replicated advances, the whole scientific ecosystem (from authors to reviewers to editors and publishers) must evolve to embrace and promote adoption of these practices. The Information Sharing Statement and the detailed content that can be included therein, is a small but important step in this continuous process to improve the impact of the neuroinformatics resources that drive neuroscience discovery.

[7] Eickhoff, S., Nichols, T. E., Van Horn, J. D., & Turner, J. A. (2016). Sharing the wealth: Neuroimaging data repositories. *NeuroImage*, *124*(Pt B), 1065–1068. Doi:10.1016/j.neuroimage.2015.10.079

[8] Fenner, M., Crosas, M., Grethe, J., Kennedy, D., Hermjakob, H., Rocca-Serra, P., … Clark, T. (2016). A Data Citation Roadmap for Scholarly Data Repositories. *bioRxiv*, 97,196. Doi: 10.1101/097196

[9] Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R. R., Duerr, R., … Clark, T. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ. Computer Science*, *1*. Doi: 10.7717/peerj-cs.1

[10] Smith, A. M., Katz, D. S., & Niemeyer, K. E. (2016). Software citation principles. *PeerJ Computer Science*, *2*, e86. Doi: 10.7717/peerj-cs.86

[11] Bandrowski, A., Brush, M., Grethe, J. S., Haendel, M. A., Kennedy, D. N., Hill, S., … RINL Resource Identification Initiative. (2016). The Resource Identification Initiative: A Cultural Shift in Publishing. *Neuroinformatics*, *14*(2), 169–182. Doi: 10.1007/s12021–015–9284-3