EDITORIAL

Palgrave Communications' commitment to promoting transparency and reproducibility in research

Gino D'Oca[1] and Iain Hrynaszkiewicz[2]

There is increasing recognition among academics—particularly those engaged in quantitative and experimental research—that sharing of research data and reproducibility of results should be a standard part of the research culture. In the humanities and social sciences (HSS), this shift has been encouraged by prominent calls for more reproducible research in various disciplines within our scope, such as political science (http://datacommunity.icpsr.umich.edu/da-rt-workshop; Gherghina and Katsanidou, 2013), economics and sociology (Miguel *et al.*, 2014; http://www.asanet.org/images/asa/docs/pdf/CodeofEthics.pdf).

The challenge remains how to ensure that data sharing occurs effectively and consistently, particularly where it is not possible to disclose research data publicly because of privacy or sensitivity issues. In some research fields, well-established data repositories and standards already provide compelling demonstrations of the value of data sharing and reuse. A hesitancy to adopt data sharing practices in some disciplines, however, can be attributed to various factors, including: the lack of incentives and rewards for sharing research data; concerns over misleading secondary analyses; protection of those involved in sensitive or human subjects research; and a perceived lack of suitable repositories for research data (http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/).

Academic journals can play an important role in encouraging consistent and effective data sharing by implementing strong, progressive editorial policies. Evidence suggests (Wicherts *et al.*, 2006; Vines *et al.*, 2013) that journal policies on data sharing are much more effective when they are mandated, and actively checked by editors and reviewers rather than simply encouraging authors to comply. Implementing these polices takes time and effort but this is a process to which *Palgrave Communications* is committed.

After consultation with our Editorial Board we recently introduced further editorial measures and author resources to support more transparent research – which in turn should increase reproducibility and reliability. In this editorial, we provide an overview of our policies and resources, and highlight examples of good practice.

Our policy
Since its launch *Palgrave Communications*' editorial policies have included requirements for authors of original research articles to provide editors, reviewers and, wherever possible, the journal's readers with access to any research datasets and materials (http://www.palgrave-journals.com/palcomms/about/editorial-policies#Availability) on which submitted manuscripts' claims are based. We also encourage formal citation of datasets (Hrynaszkiewicz and Acuto, 2015).

The aim of our editorial policy is to publish reliable research for the benefit of the academic communities we serve. Sharing of research data has been associated with increased citations of papers in some disciplines (Piwowar and Vision, 2013), showing a potential benefit to individual authors as well as the wider community. Furthermore, we are mindful of ensuring our readers have a consistent experience and that they are able to make informed assessments when reading research articles that are supported by original research data.

'Data Availability' statements
After consultation with our large, multidisciplinary Editorial Board, we have further strengthened our editorial policy to provide more transparency and consistency about data availability for our readers: all original articles will now have to include a 'Data Availability' statement, and, wherever possible, a link to and citation of datasets when these are available in an appropriate public repository.

We recognize it is not always possible to share research data publicly, such as when individual privacy could be compromised,

[1] Managing Editor, Nature Publishing Group & Palgrave Macmillan, The Macmillan Campus, Trematon Walk, Wharfdale Road, London, N1 9FN, UK [2] Head of Data and HSS Publishing, Open Research, Nature Publishing Group & Palgrave Macmillan, The Macmillan Campus, Trematon Walk, Wharfdale Road, London, N1 9FN, UK Correspondence: (e-mail: Palcomms@palgrave.com)

and in such instances data availability will still be stated in the manuscript along with any conditions for access. The journal's editorial team carries out appropriate checks to ascertain whether non-availability of data is for legitimate reasons.

'Data Availability' statements will take one of the following forms (or a combination of more than one if required for multiple datasets):

1. The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
2. The datasets generated during and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.
3. Data sharing not applicable to this article as no datasets were generated or analysed during the current study.
4. All data generated or analysed during this study are included in this published article.

While deposition of data in specialized public repositories (statement #1) is the preferred approach wherever possible, in principle all our authors who generate or analyse research data will be expected to comply with #2, which is the minimum journal policy. We check compliance with this policy at an early stage of the peer-review process. We have also updated our peer reviewer guidelines to ask reviewers to assess authors' compliance with our policy and the expectations of the relevant research community (http://www.palgrave-journals.com/palcomms/referees#writing).

Pre-registration of studies and code sharing

We also recognize the importance of the pre-registration of studies and code sharing as ways of making research more citable, discoverable, interpretable, reusable and reproducible. Our editorial policy guidelines now encourage pre-registration of studies where appropriate databases exist (for example, www.socialscienceregistry.org; http://opensscienceframework.org/; http://egap.org/design-registration/; http://ridie.3ieimpact.org/). We also request that, when applicable, authors make available, for the benefit of editors and reviewers, any previously unreported custom computer code used to generate results that are reported in a paper and which are central to its main claims. On publication, authors will be encouraged to release custom computer code in a way that allows readers to repeat the published results, ideally using a public repository that can assign it a Digital Object Identifier (DOI; a persistent web link).

Our own repository

Data repositories are generally the most appropriate place for research data as they provide independent discoverability, permanence and consistent metadata for datasets. We recommend that authors who make their datasets and code publicly available use the most broadly supported and recognized repository for their research community. We support authors so they can comply with our data policies, and get the most benefit from them; for instance, we provide a list of trusted repositories, which will evolve over time (http://www.palgrave-journals.com/palcomms/about/editorial-policies#Availability).

We recognize that some authors, however, do not have a preferred repository. In collaboration with Dataverse at Harvard, we now provide a dedicated repository for *Palgrave Communications* (https://dataverse.harvard.edu/dataverse/palcomms). This Dataverse allows the confidential (non-public) submission and peer review of datasets before publication, and their subsequent open release with published articles. Dataverse has established itself as a preferred repository for many social HSS journals and disciplines, and provides valuable features such as DOIs for published datasets enabling them to be formally cited in reference lists.

Examples of good practice

Replication data for a recent Palgrave Communications article
A research paper recently published in *Palgrave Communications*, by Shutters et al. (2015a), provides an example of effective data sharing. The authors analyse data from 400 US metropolitan statistical areas, and set out to quantify the capacity of a complex system (in this case an urban economy) to absorb disturbance and reorganize, with the aim of predicting responses to major shocks like the recent global recession. All datasets analysed in the study, which were derived from public domain governmental resources, are available in the journal's Dataverse (Shutters et al., 2015b) and the manuscript includes a data availability statement.

Anonymous survey data from our Editorial Board consultation
Before rolling out the policy changes described, we carried out a consultation with our Editorial Board. *Palgrave Communications* is broad in scope, encompassing fields of study, such as economics, complex networks and psychology, to which our enhanced data sharing policy is immediately relevant, as well as other disciplines whose output is typically more theoretical or qualitative, and for which data reproducibility is therefore of less pressing concern. All participants were asked a range of questions regarding their own data sharing practices and preferences.

Feedback was obtained from researchers covering the full spectrum of disciplines covered by the journal. Over half of the 49 respondents (51.1%) stated that, at least some of the time, their research generates numerical or other types of data files that would be subject to the journal's new data sharing policy; and approximately 40% use data generated by another researcher (obtained by direct personal contact or from a repository) in their own research. Approximately one-third of respondents stated they would be likely to deposit data in *Palgrave Communications*' Dataverse following submission. A total of 55.3% agreed that sharing of research data benefits their own research; only 2.1% disagreed. As the anonymized survey data are available in Dataverse (Palgrave, 2015), readers are free to explore, share and reuse the data. By default, authors depositing in Dataverse make their data available under the Creative Commons CC0 public domain dedication, enabling maximum potential for reuse (http://creativecommons.org/about/cc0).

The survey invited participants to share additional comments. Concerns about data sharing tended to relate to issues surrounding anonymity, consent and ownership, and also the appropriateness of sharing certain datasets, such as those whose collection has been costly or complex, or data that are very specific to a particular niche area of research. Our sample size is very small but the findings echo other, larger surveys (http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/). Several others said that datasets come in many forms and, as such, any editorial policy would need to judge each on a case-by-case basis. By offering authors the option to select and adapt various types of data availability statement, we are satisfied that such nuances can be addressed. We will monitor feedback from authors and our Editorial Board, and consider evolving our policies in the future.

We will continue to work with our authors to promote the publication of reproducible research. In addition, we are very interested to hear from researchers or groups who are actively engaged in encouraging such practice in HSS and business research.

References

Gherghina S and Katsanidou A (2013) A data availability in political science journals. *European Political Science*; **12**, 333–349, doi:10.1057/eps.2013.8.

Hrynaszkiewicz I and Acuto M (2015) *Palgrave Communications*—Connecting research in the humanities, social sciences and business. *Palgrave Communications*; **1**, 1400610.1057/palcomms.2014.6.

Miguel E *et al.* (2014) Social science: Promoting transparency in social science research. *Science*; **343** (6166): 30–31.

Palgrave (2015) *Palgrave Communications*' editorial board data sharing survey—Raw data. Dataverse, http://dx.doi.org/10.7910/DVN/JVZHFG, accessed 1 May.

Piwowar H A and Vision T J (2013) Data reuse and the open data citation advantage. *PeerJ* 1:e175 https://dx.doi.org/10.7717/peerj.175.

Shutters S T, Muneepeerakul R and Lobo J (2015a) Quantifying urban economic resilience through labour force interdependence. *Palgrave Communications*; **1**, 15010, doi:10.1057/palcomms.2015.10.

Shutters S, Muneepeerakul R and Lobo J (2015b) Replication data for: Quantifying urban economic resilience through labour force interdependence. Dataverse, http://dx.doi.org/10.7910/DVN/HEWGDD, accessed 1 May.

Vines T H *et al.* (2013) Mandated data archiving greatly improves access to research data. *FASEB Journal*; **27** (4): 1304–1308.

Wicherts J M, Borsboom D, Kats J and Molenaar D (2006) The poor availability of psychological research data for reanalysis. *American Psychologist*; **61** (7): 726–728.

Data Availability

The dataset generated from the anonymous survey conducted with Editorial Board members is publicly available in the Dataverse repository (Palgrave, 2015): http://dx.doi.org/10.7910/DVN/JVZHFG

Additional Information