brazilianpoliticalsciencereview

REVIEW ESSAY

Seven Reasons Why: A User's Guide to Transparency and Reproducibility*

Dalson Figueiredo Filho**
https://orcid.org/0000-0001-6982-2262
**Pernambuco Federal University, Recife, Pernambuco, Brazil
Rodrigo Lins**
https://orcid.org/0000-0002-4521-3806
**Pernambuco Federal University, Recife, Pernambuco, Brazil
Amanda Domingos**
https://orcid.org/0000-0002-2900-5337
**Pernambuco Federal University, Recife, Pernambuco, Brazil
Nicole Janz***
https://orcid.org/0000-0002-1668-166X
***University of Nottingham, Nottingham, United Kingdom
Lucas Silva**
https://orcid.org/0000-0002-5013-6278
**Pernambuco Federal University, Recife, Pernambuco, Brazil

Despite a widespread agreement on the importance of transparency in science, a growing body of evidence suggests that both the natural and the social sciences are facing a reproducibility crisis. In this paper, we present seven reasons why journals and authors should implement — transparent guidelines. We argue that sharing replication materials, which include full disclosure of the methods used to collect and analyze data, the public availability of raw and manipulated data, in addition to computational scripts, may generate the following positive outcomes: 01. production of trustworthy empirical results, by preventing intentional frauds and avoiding honest mistakes; 02. making the writing and publishing of papers more efficient; 03. enhancing the reviewers' ability to provide better evaluations; 04. enabling the continuity of academic work; 05. developing scientific reputation; 06. helping to learn data analysis; and 07. increasing the impact of scholarly work. In addition, we review the most recent computational tools to work reproducibly. With this paper, we hope to foster transparency within the political science scholarly community.

Keywords: Transparency; reproducibility; replication.

*http://dx.doi.org/10.1590/1981-3821201900020001
For data replication, see: https://doi.org/10.7910/DVN/BYDCCA. This publication is registered under a CC-BY Licence.
This work was supported by the British Academy / Newton Mobility Fund, Berkeley Initiative for Transparency in the Social Sciences (BITSS), Project Teaching Integrity in Empirical Research (TIER), Fundo de Amparo à Ciência e Tecnologia do Estado de Pernambuco and the University of Nottingham's research area 'Rights and Justice'.

The scholarly interest in transparency, reproducibility and replication has been rising recently, with compelling evidence that many scientific studies have failed to be replicated (GOODMAN, FANELLI and IOANNIDIS, 2016). In political science, Key (2016) reports that 67.6% of all articles published in the American Political Science Review (APSR) between 2013 and 2014 do not provide replication materials. Similar trends are found in sociology (LUCAS et al., 2013), economics (CHANG and LI, 2015; DEWALD et al., 1986; KRAWCZYK and REUBEN, 2012), international relations (GLEDITSCH and JANZ, 2016), psychology (NOSEK et al., 2015), medicine (SAVAGE and VICKERS, 2009), biostatistics (LEEK and PENG, 2015) and genetics (MARKOWETZ, 2015). In short, both the natural and the social sciences are facing a reproducibility crisis (FREESE and PETERSON, 2017; LEEK and PENG, 2015; MUNAFÒ et al., 2017).

To make the problem even worse, there is no agreement as to what transparency, reproducibility and replication mean exactly (JANZ, 2016). Scholars from different fields use these concepts without a common understanding of their definition (PATIL, PENG and LEEK, 2016, p. 01). For example, King (1995) argues that "the replication standard holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party could replicate the results without any additional information from the author" (KING, 1995, p. 444). Jasny et al. (2011) define replication as "the confirmation of results and conclusions from one study obtained independently in another" (JASNY et al., 2011, p. 1225). According to Hamermesh (2007), replication has three perspectives: pure, statistical and scientific. Seawright and Collier (2004) argue that "two different research practices are both called replication: a narrow version, which involves reanalyzing the original data, and a broader version based on collecting and analyzing new data" (SEAWRIGHT and COLLIER, 2004, p. 303). The plurality of definitions goes on[1].

[1]For example, "all data and analyses should, insofar as possible, be replicable (...) only by reporting the study in sufficient detail so that it can be replicated is it possible to evaluate the procedures followed and methods used" (KING, KEOHANE and VERBA, 1994, p. 26). Herrnson (1995) argues that replication, verification and reanalysis have different meanings. In particular, he states that "replication repeats an empirical study in its entirety, including independent data collection (...) replication increases the amount of information for an empirical research question and increases the level of confidence for a set of empirical generalizations" (HERRNSON, 1995, p. 452).

The meaning of reproducibility is also mutable. Goodman, Fanelli and Ioannidis (2016) argue that the current use of reproducible research was initially employed in computational science not for corroboration but transparency. They also propose a new lexicon for research reproducibility in three dimensions: methods, results and inferences. The U.S. National Science Foundation (NSF) defines reproducibility as "the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator" (NSF, 2015, p. 06). Regarding transparency, the terminology is also ambiguous. For example: according to Janz (2016), "working transparently involves maintaining detailed logs of data collection and variable transformations as well as of the analysis itself" (JANZ, 2016, p. 02). Moravcsik (2014) defines transparency as "the principle that every political scientist should make the essential components of his or her work visible to fellow scholars" (MORAVCSIK, 2014, p. 48). He also identifies three dimensions of transparency: data, analytic and production. Although these conceptual disagreements seem to be mainly semantic, they may also have scientific and policy implications (PATIL, PENG and LEEK, 2016). Therefore, to avoid conceptual misunderstandings, we adopt the following definitions:

Table 01. Replication, reproducibility and transparency

Replication	Reproducibility	Transparency
Process by which a published article's findings are re-analyzed to confirm, advance or challenge the original results (JANZ, 2016).	Set of computational functions (scripts) that allows to exactly reproduce the observed results from raw data (PATIL, PENG and LEEK, 2016).	The full disclosure of the research design, which includes the methods used to collect and analyze data, the public availability of both raw and manipulated data, in addition to the computational scripts employed along the way.

Source: Elaborated by the authors.

Replication research, following Gary King's definition (1995), involves often as a first step using the same data with the same statistical tools (this first step aims at verification of research could also be called 'duplication'). However, a full replication also collects new data to test the same hypothesis, or includes a new variable to test the same model. More broadly, we may define replication as the process by which a published article's findings are re-analyzed to confirm, advance or challenge the robustness of the original results (JANZ, 2016). The American Political Science Association (APSA) guidelines highlight the following

features meant to enhance reproducibility: 01. data access; 02. data gathering description and 03. details on data transformation and analysis. In this paper, the concept of reproducibility is closely related to computational functions that allow the researcher to reproduce the reported results exactly using the raw data (PATIL, PENG and LEEK, 2016).

Lastly, we should define transparency. In physics, the concept of transparency refers to objects that allow the transmission of light without appreciable scattering so that bodies lying beyond are seen clearly. Different from translucent and opaque, transparent ways produce regular and well defined trajectories (SCHNACKENBERG, 2009). In scientific parlance, transparency means full disclosure of the process by which the data are generated and analyzed. According to King, Keohane and Verba (1994), "without this information we cannot determine whether using standard procedures in analyzing the data will produce biased inferences. Only by knowing the process by which the data were generated will we be able to produce valid descriptive or causal inferences" (KING, KEOHANE and VERBA, 1994, p. 23). Miguel et al. (2014) argue that transparent social science covers three core practices: "disclosure, registration and pre-analysis plans, and open data and materials" (MIGUEL et al., 2014, p. 30). Following this reasoning, we define transparency as the full disclosure of the research design, which includes the methods used to collect and analyze data, the public availability of both raw and manipulated data, in addition to the computational scripts employed along the way[2]. From now on, we should employ replication materials as an empirical indicator to represent the concept of transparency.

In this paper, we present seven reasons why journals and authors should adopt transparent guidelines. We argue that sharing replication materials, which include full disclosure of the methods used to collect and analyze data, the public availability of both raw and manipulated data, in addition to the computational scripts that were used, may generate the following positive outcomes: 01.

[2]According to King, Keohane and Verba (1994), "if the method and logic of a researcher's observations and inferences are left implicit, the scholarly community has no way of judging the validity of what was done. We cannot evaluate the principles of selection that were used to record observations, the ways in which observations were processed, and the logic by which conclusions were drawn. We cannot learn from their methods or replicate their results. Such research is not a 'public' act. Whether or not it makes good reading, it is not a contribution to social science" (KING, KEOHANE and VERBA, 1994, p. 08).

production of trustworthy empirical results, by preventing intentional frauds and avoiding honest mistakes; 02. making the writing and publishing of papers more efficient; 03. enhancing the reviewers' ability to provide better evaluations; 04. enabling the continuity of academic work; 05. developing scientific reputation; 06. helping to learn data analysis; and 07. increasing the impact of scholarly work.

Also, we review the most recent computational tools to work reproducibly. We are aware that this research note is no substitute for a careful reading of primary sources and materials of a more technical nature. However, we believe that scholars with no background on the subject will benefit from a document having predominantly pedagogical goals[3].

To the best of our knowledge, this paper provides the first attempt to bring both transparency and reproducibility to the Brazilian political science research agenda[4]. Generally speaking, scholars in the United States and Europe produce the majority of work on the topic. Now is the time to expand scientific openness beyond the mainstream of our discipline. Also, we believe that we are on the right track, after the program of the XI Meeting of the Brazilian Political Science Association (ABCP)[5] was made public. For the first time, the program had a session

[3]Although our recommendations are focused on quantitative, variance-based methods, there is a growing literature on transparency in qualitative research. For instance, Moravcsick (2014) examines the main challenges posed for qualitative scholars. Elman and Kapiszewski (2014) evaluate if and how qualitative pundits can disclose more about the processes through which they generate and analyze data. To this end, the American Political Science Association (APSA) has produced the Guidelines for Data Access and Research Transparency for Qualitative Research in Political Science. In addition, the Qualitative Transparency Deliberations (<https://www.qualtd.net/>) "was launched in 2016 to provide an inclusive process for deliberation over the meaning, costs, benefits, and practicalities of research transparency, openness in qualitative political science empirical research". We are thankful to the referee of the BPSR on this matter. See also <https://www.qualtd.net/page/resources>.

[4]We reviewed all articles published in four top Brazilian national journals between 2010 and 2017 and found no publication that deals with replication, reproducibility and transparency. We reviewed articles published in the following journals: 01. DADOS; 02. Brazilian Political Science Review; 03. Revista de Sociologia e Política and 04. Opinião Pública using 'replication', 'transparency' and 'reproducibility' as keywords in both Scielo and Google Scholar. As far as we are concerned, there is only one Brazilian political science journal that has a strict policy of data sharing before publishing the papers, namely, the Brazilian Political Science Review. See <http://bpsr.org.br/files/archives/database.html>. According to our keywords search, it seems that Galvão, Silva and Garcia (2016) produced the only paper on the subject published in a Brazilian journal. We also found a blog post from Scielo which is available at <http://blog.scielo.org/en/2016/03/31/reproducibility-in-research-results-the-challenges-of-attributing-reliability/#.Wn7IOudG02w>.

[5]<https://cienciapolitica.org.br/sites/default/files/documentos/2018/07/programacao-xi-encontro-abcp-2018-1371.pdf>

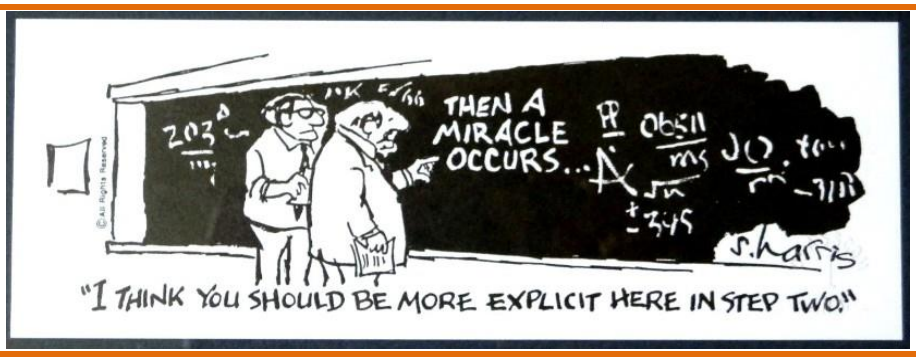
on Replication and Transparency led by professors Lorena Barberia (USP) and Marcelo Valença (UERJ). George Avelino and Scott Desposato, in particular, presented a paper on the subject focusing on the Brazilian case.

The remainder of the article consists of four sections: the one following this, which explains the benefits of creating replication materials; then we describe some of the tools for reproducible computational research; we proceed by examining different sources where one can learn about transparent investigation, and we conclude by suggesting what could be done to foster openness within the political science scholarly community.

Seven reasons to require replication materials[6]

Figure 01 shows what journal editors and authors should avoid. Transparent research requires full disclosure of the procedures used to obtain and analyze data as pointed out by the Data Access and Research Transparency protocol[7]. The higher the transparency, the easier it is to replicate and reproduce the results. In what follows, we review Markowetz's (2015) reasoning on the importance of transparency as a standard scientific practice and we add two more motives why scientific journals should require reproducible materials, and why authors should follow these guidelines.

Figure 01. The miracle of Science



Source: <https://shadowtolight.wordpress.com/2015/03/21/defeating-jerry-coynes-argument-about-science-and-religion-part-2/>. Accessed on July, 07, 2018.

[6]We are in debt to the BPSR referee who pointed out that our focus should be on journals' policies rather than on author ethos. According to her review, "it is not a problem of researchers' goodwill, but a matter of replication policies adopted by the journals, which are the gatekeepers of publications".

[7]See <https://www.dartstatement.org/>.

Reason 01. Replication materials help to avoid disaster

Scientific frauds are underestimated (FANG et al., 2012). From medicine to physics, from chemistry to biology, it is not hard to find cases of scientific misconduct (FANELLI, 2009). For example, Baggerly and Coombes (2009) demonstrate how the results reported by Potti et al. (2006) vanish after some spreadsheet problems are fixed. Also, they provide evidence of data fabrication in the original study. As Young and Janz (2015) argued, "a veritable firestorm hit political science" (YOUNG and JANZ, 2015) when the LaCour and Green paper was retracted by Science due to the alleged use of fictional data. At the time, LaCour was a PhD student at the University of California at Los Angeles (UCLA). Using a field experiment, they reported that conversations with gay canvassers produce significant changes in voters' opinions on gay marriage[8]. The paper was very influential but failed to be reproduced by two doctoral students from the University of California at Berkley because, as it turned out later, the data was fabricated[9]. The LaCour case was a wakeup call for political scientists against scientific misconduct. According to Simonsohn (2013), "if journals, granting agencies, universities, or other entities overseeing research promoted or required data posting, it seems inevitable that fraud would be reduced" (SIMONSOHN, 2013, p. 1875)[10]. Thus, reproducible mandatory policies on reproducibility can reduce the likelihood of scientific frauds.

Besides preventing intentional deception, reproducible research also avoids honest mistakes. Again, there are many examples of how minor errors can jeopardize an entire research enterprise. One of the most well-known relates to Reinhart and Rogoff's paper (2010), where a small error in the data Excel spreadsheet compromised the empirical results. As King (1995) noted more than 20 years ago "the only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis

[8]More detailed description of the case can be found here: <http://www.chronicle.com/article/What-Social-Science-Can-Learn/230645/>.
[9]See: <http://www.chronicle.com/article/We-Need-to-Take-a-Look-at/230313/>.
[10]This is not to say that reproducibility can always avoid scientific misconduct. Researchers can fabricate data and use the correct codes, but it is also possible to create fake data with forged scripts (which hopefully is not a widespread practice). We are thankful to BPSR reviewers on this specific matter.

produced" (KING, 1995, p. 444). In 2016, Imai, King and Rivera found discrepant data in the work of Ana de La O (2013; 2015). In both of her works, De La O found a partisan effect of a nonpartisan programmatic policy. However, Imai, King and Rivera (2016) showed that her dataset contained data observations that are likely to be labeled as outliers. Figure 02 reproduces this information.

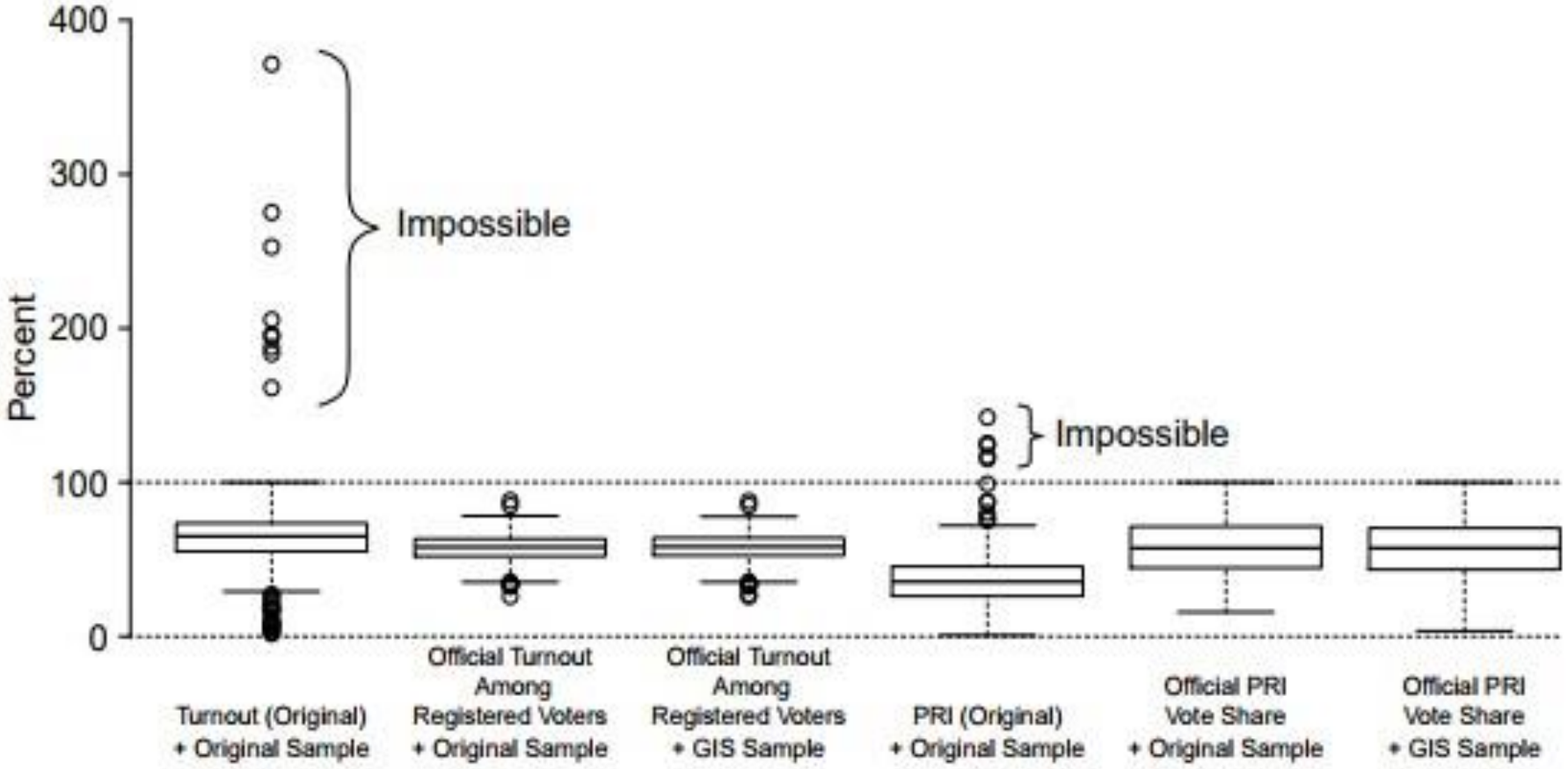
For example, the turnout variable only varies from 0 to 100. When Imai, King and Rivera (2016) corrected the data, the reported effects vanished. If the dataset had been publicly available from the start, any student with minimal statistical training would have been able to spot Ana de La O's mistakes. Similarly, if all journals adopted a strict reproducibility policy that required both data and code, the reviewers would probably be the first to spot the mistake. Therefore, the first reason why journal editors should require reproducible materials is to prevent frauds and reduce mistakes.

Reason 02. Transparency makes it easier to write papers

Journals should adopt transparency guidelines because it facilitates paper writing for their authors. In political science, King (1995, 2006) initially emphasized this rationale. Creating replication materials will force the researcher to plan every step of their work. This will help in having a clearer idea of the research design, as well as knowing exactly where the data is stored and how it was managed. This would help at the time of the research, but also for future inquiries. On the aggregate, better individual papers would lead to a more robust collective scientific endeavour (MARKOWETZ, 2015).

Ball and Medeiros (2012) explain that creating replication materials will make the researcher readily aware of potential problems and, consequently, submit more robust research to journals. They argue that when young scholars know that they will need to make their data management public, "their data management and analysis tend to be much more organized and efficient, and their understanding of what they are doing tends to be much greater than when they use their statistical software to execute commands interactively" (BALL and MEDEIROS, 2012, p. 188).

Figure 02. Reproducibility and outlier issues



Source: Imai et al. (2016).

With this in mind, they have created the TIER protocol, which is an applied tool for transparent research[11].

We also adopted replication exercises recently in our graduate (PhD) courses and have observed exciting results. Students are more involved and show higher quality final papers. For example, in 2015, two of our students started a replication paper project that went on to publication in a Qualis A2 journal[12]. The same effect was reported by King (2006).

Replication also enhances researchers' writing. The reasoning goes as follows: a published paper has already gone through peer evaluation, which means that the reviewers have already considered it publishable once (KING, 2006). A replication paper can focus on 01. extending the dataset (cross-section, time-series or both); 02. applying new methods to the same data; 03. controlling for new variables; and 04. investigating an entirely new research question with the same data, to name a few possibilities.

At first one can argue that replication papers are harder to publish because they lack the sense of novelty that is attractive to publications. However, there are many examples of successful replications that were published in major journals. For instance: Fraga and Hersh (2010) published in the Quarterly Journal of Political Science by replicating data from Gomez et al. (2007); Bell and Miller (2013) published in the Journal of Conflicts Resolution by replicating data from Rauchhaus (2009); and Dai (2002) published in the American Political Science Review by reexamining data from Mansfield, Milner, and Rosendorff (2000).

Reason 03. Replication materials can lead to better paper reviews

After months (sometimes years) of work, it is very disappointing to get a poor paper review. Some referees reject the paper or provide assessments that are too general and thus unlikely to improve our draft. According to the Center for Scientific Review of the National Institutes of Health (NIH), one of the significant roles of the reviewers is to ensure that the scientific foundation of the project – including reviewing the data – is sound. The document states that the goal of a

[11]Information on the TIER Protocol can be found at <http://www.projecttier.org/tier-protocol/>.
[12]See Montenegro and Mesquita (2017) at: <http://www.bpsr.org.br/index.php/bpsr/article/view/304>.

rigorous review is to "support the highest quality science, public accountability, and social responsibility in the conduct of scientific research" (NIH, 2018, p. 02). Moreover, the NIH (2018) expects the reviewers to pay attention to relevant variables in the area of research.

Markowetz (2015) argues that reproducibility is vital to increase reviewers' contribution to our work. We agree with him. Also, in order to suggest a new control variable or a more appropriate model specification, reviewers can work with the data themselves. Reviewers also can contribute to our coding by suggesting better scripts. We are not so naïve as to expect a full commitment of all reviewers in re-analyzing our data, but if both data and code are available from the start, it is easier to get better peer evaluations. For example, the journal Political Science Research and Methods (PSRM) and the American Journal of Political Science (AJPS) have a strict replication policy[13]. To have a paper accepted, authors must share replication materials in advance (data and code). The journal's staff then needs to rerun the analysis and get the same results (tables and figures). Only then do they allow the paper's publication. If, for any reason, the results are not reproducible, the author must update data and codes. In this process, reviewers also evaluate the quality of replication materials, which is likely to increase their overall contribution to enhancing the paper. Therefore, when journals require the submission of replication materials already at the peer-review stage, they can improve the quality of such reviews.

Reason 04. Replication materials enable the continuity of academic work

When journals adopt transparency guidelines, they support continuity of academic work for their authors. During the 2014 Berkeley Initiative for Transparency in the Social Sciences (BITSS) workshop, professor Ted Miguel asked who had had problems when trying to replicate their own papers. People slowly started to raise their hands, and we found ourselves in an embarrassing situation: we were all attending a transparency meeting, but most of us did not have a reproducible research culture. Markowetz (2015) points out that we can trace personal reproducibility problems by "documenting data and code well and making

[13]We are thankful to the BPSR referee for this warning.

them easily accessible" (MARKOWETZ, 2015, p. 03). According to King (1995), "without adequate documentation, scholars often have trouble replicating their own results months later" (KING, 1995, p. 444). Finifter (1975) developed a taxonomy of nine replication strategies "to clarify and codify how replications contribute to confidence and cumulative advances in substantive research" (FINIFTER, 1975, p. 120). Goodman et al. (2015) offer a short guide on how to increase the quality of scientific data. Specifically, they propose ten rules to facilitate data reuse[14]. In 2016, Science published the following Editorial expression of concern:

> In the 03 June issue, *Science* published the Report "Environmentally relevant concentrations of microplastic particles influence larval fish ecology" by Oona M. Lönnstedt and Peter Eklöv. The authors have notified *Science* of the theft of the computer on which the raw data for the paper was stored. These data were not backed up on any other device nor deposited in an appropriate repository. *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce or extend the conclusions of the paper" (BERG, 2016, p. 1242).

An unfortunate example shows how reproducibility enables continuity of academic work. In September 2017, Noxolo Ntusi fought against two thieves to protect her master's thesis. She explained that all the relevant information for her research was stored in an HD drive in her bag[15]. Likewise, in November 2017, a PhD student offered a $ 5,000 reward for his data stolen in Montreal. He said, "I have a backup of the raw data left on equipment in the lab, but I do not have a backup of the data analysis that I have been doing for the last few years" (MONTREAL GAZETTE, 2017). In our own personal experience, we have observed a similar case - a student who used to keep the only copy of her dissertation in a 1.44 floppy disk. At some point, the file got corrupted, and she lost all her work. We can easily avoid

[14]Rules: "01. Love your data and help others love it too; 02. Share your data online, with a permanent identifier; 03. Conduct science with a particular level of reuse in mind; 04. Publish workflow as context; 05. Link your data to your publications as often as possible; 06. Publish your code (even small bits); 07. State how you want to get credit; 08. Foster and use data repositories; 09. Reward colleagues who share their data properly and 10. Be a booster for data science" (GOODMAN et al., 2015).

[15]<http://www.bbc.com/portuguese/internacional-41261885> Video of her fighting is available at <https://www.youtube.com/watch?v=gj7bhJpYEqM>.

these troubles by creating replication materials and storing data, codes and other relevant information in online repositories.

Unfortunately, we should not expect a natural behavioral change from authors. We should push for a change in scientific editorial policies, which would be more likely to induce higher compliance. According to Stockemer, Koehler and Lenz (2018), "if we as a discipline want to abide by the principle of research and data transparency, then mandatory data sharing and replication are necessary because many authors are still unwilling to share their data voluntarily or make unusable replication materials available" (STOCKEMER, KOEHLER and LENZ, 2018, p. 04).

Reason 05. Replication materials help to build scientific reputation

Ebersole, Axt and Nosek (2016) conducted a survey of adults (N = 4,786), undergraduates (N = 428), and researchers (N = 313) and found that "respondents evaluated the scientist who produces boring but certain (or reproducible) results more favorably on almost every dimension compared to the scientist who produces exciting but uncertain (or not reproducible) results" (EBERSOLE, AXT and NOSEK, 2016, p. 07). Although transparency is key to confidence in science, most political science academic journals do not adopt strict replication policies. Gherghina and Katsanidou (2013) examined 120 international peer-reviewed political science journals and found out that only 18 (15%) had a replication policy. In Brazil, the Brazilian Political Science Review (BPSR) is the only one that has a data sharing mandatory policy. This institutional decision represents a significant advance for Brazilian editors, and other journals should emulate this decision shortly. Internationally, both the American Journal of Political Science (AJPS) and American Political Science Review (APSR) have recently adopted transparent data sharing policies[16].

Also, following Markowetz's (2015) remarks, if you ever have a problem with your data, "you will be in a very good position to defend yourself and to show that you reported everything in good faith" (MARKOWETZ, 2015, p. 04). For

[16]We should note that BPSR, AJPS and APSR have different replication policies. While the BPSR only requires data sharing, AJPS requires both data and code which is then replicated by journal staff members. It is important to share both data and code. We are thankful to the BPSR referee for this warning.

example, in 2009 the U.S. Court of Appeals upheld a district court decision to dismiss John R. Lott's claims against Steven Levitt. The book Freakonomics mentioned Lott's name, and he filed a defamation suit against Levitt and HarperCollins, the publisher. In this case, Levitt argued that Lott's results were not reproducible and the plaintiff understood this passage as an accusation of academic dishonesty[17]. Time, energy and resources could have been saved if Lott's replication materials had been made publicly available.

In short, transparency adds two gains to the reputation of journals as well as authors: the readers of your work will know that you took the time to make all your documentation available, thus giving it more credibility, and you are shielded and more likely to be protected from any misconduct charges, given that you did everything in good faith (and honest errors are human). Again, we should not rely solely on the authors' goodwill. Journal editors play a crucial role in this process, since they can change the incentives and induce open research practices by adopting mandatory editorial replication policies. For journals, the reputational benefit is clear: they can avoid negative publicity when an article published by them fails to replicate, must be corrected or even retracted.

Reason 06. Replication materials help to learn data analysis

King (1995) initially stressed this reasoning: "reproducing and then extending high-quality existing research is also an extremely useful pedagogical tool" (KING, 1995, p. 445). According to Janz (2016), replication provides students with "a better way to learn statistics: Replication is essential to a deeper understanding of statistical tests and modeling. The advantage over textbook exercises is that students use real-life data with all bugs and complications included" (JANZ, 2016, p. 05). From our teaching experience, we have observed that students are more motivated to work with real data and arriving at the same figures than working with some dull, repetitive homework assignment unrelated to what they are studying. Creating replication materials can foster data analysis learning by providing examples for the students. For instance, Gujarati (2011)

[17]See: <http://caselaw.findlaw.com/us-7th-circuit/1139405.html>.

published a book called 'Econometrics by Example', where he teaches each topic by showing applied cases. It is easier for the students to learn from examples related to their fields of study than from areas that are alien to them.

Suppose that you are learning about data transformation. The primary purpose of your statistics professor is to teach you how to improve the interpretability of graphs when dealing with skewed data (see Figure 03). He shares with you a dataset on animals that have only two variables (body weight and brain weight) and asks what the best mathematical function to be applied in order to get a normal distribution is.

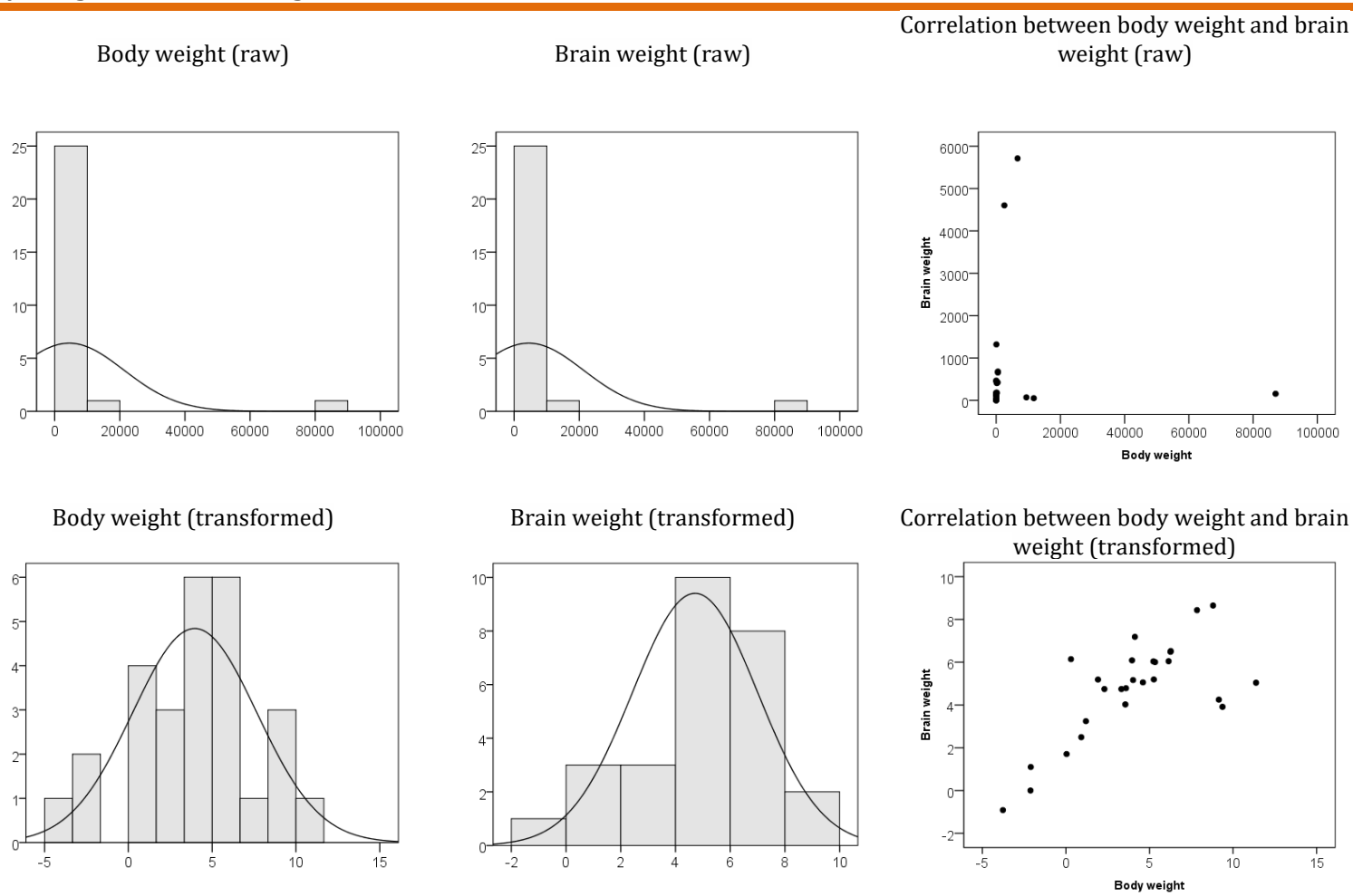
After some discussion, you learn that the natural logarithmic transformation is the right answer. Now imagine that the same professor handles your data on the relationship between campaign spending and electoral outcomes. In which scenario are you more likely to engage in active learning? On average, we would expect political science students to be more interested in money that flows to the political system than on animals' biological features.

Following Janz's (2016) reasoning, we believe that professors can foster a transparent culture by adopting exercises into their academic activities. How exactly are learning purposes related to journal editorial policies? We argue the following: First of all, students can learn from replicating published work much easier when the authors provide their materials. Second, if journals required mandatory replication materials, both students and experienced scholars would have to change their behavior when thinking about publishing. Students, particularly, would develop higher data analysis skills that are likely to improve the overall quality of paper submissions. In the end, journal editors would receive higher quality submissions and scientific endeavor would thrive.

Reason 07. Replication materials increase the impact of scholarly work

There are few events more frustrating in academic life than being ignored by the profession, and empirical evidence suggests that "the modal number of

Figure 03. Body weight and brain weight



Source: Elaborated by the authors from <https://mste.illinois.edu/malcz/DATA/BIOLOGY/Animals.html>.

citations to articles in political science is zero: 90.1% of all articles are never cited" (KING, 1995, p. 445). In other words, we write for nobody. We attend academic conferences, we apply for both internal and international grants, and we revise and resubmit the same paper as much as necessary in order to get it published in a high impact journal. Why do we invest time, resources and pride to get our message out there? Right answer: to advance scientific knowledge.

We argue that transparency is a crucial resource to increase the impact of scholarly work. According to Gleditsch, Metelits and Strand (2003), papers that share data are twice more cited compared to those that do not. Similarly, Piwowar, Day and Fridsma (2007) reported that public data availability is associated with a 69% increase in citations, controlling for other variables. According to King (1995), "an article that cannot be replicated will generally be read less often, cited less frequently, and researched less thoroughly by other scholars" (KING 1995, p. 445). Ghergina and Katsanidou (2013) found a positive association between impact factor and the likelihood of adopting a transparent editorial policy, after controlling for age of the journal, frequency, language and type of audience. More recently, Strand et al. (2016) estimated an unconditional fixed effects negative binomial regression model with a sample of 430 articles from the Journal of Peace Research. They found that sharing data increases citations, even controlling for scholars' name recognition.

According to a Nature editorial on transparency in science, "the benefits of sharing data, not only for scientific progress but also for the careers of individuals, are slowly being recognized" (NATURE GEOSCIENCE, 2014, p. 777). Table 02 summarizes the guidelines of the Transparency and Openness Promotion (TOP).

Following academic literature, it seems that Level 03 citation standards would lead to higher academic impact. Two interesting examples illustrate how data sharing can enhance the impact of scientific work. The first comes from Paasha Mahdavi. According to his Google Scholar profile, the most cited work of his career is not a paper but a dataset[18]. Overall, 48.10% of all his citations are concentrated in Oil and Gas Data (1932-2014). Similarly, John M. Powell designed his most influential paper to disseminate a time series cross-section dataset on military

[18]Ross and Mahdavi (2015) share data on Oil and Gas (1933-2014) and got 86 citations. See: <https://scholar.google.com/citations?user=OmTdT0MAAAAJ&hl=en>.

Table 02. Summary of the eight standards and three levels of the TOP guidelines

	Level 0	Level 01	Level 02	Level 03
Citation standards	Journal encourages citation of data, code, and materials – or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing – or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication
Analytic methods (code) transparency	Journal encourages code sharing – or says nothing.	Article states whether code is available and, if so, where to access it.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication
Research materials transparency	Journal encourages materials sharing – or says nothing	Article states whether materials are available and, if so, where to access them	Materials must be posted to a trusted repository. Exceptions must be identified at article submission	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication

to be continued

continuation

	Level 0	Level 01	Level 02	Level 03
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates Design transparency standards	Journal requires adherence to design transparency standards for review and publication	Journal requires and enforces adherence to design transparency standards for review and publication
Pre registration of studies	Journal says nothing	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements	Journal requires preregistration of studies and provides link and badge in article to meeting requirements
Pre registration of analysis plans	Journal says nothing.	Journal encourages pre-analysis plans and provides link in article to registered analysis plan if it exists	Journal encourages pre-analysis plans and provides link in article and certification of meeting registered analysis plan badge requirements	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies – or says nothing	Journal encourages submission of replication studies	Journal encourages submission of replication studies and conducts blind review of results	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes

Source: Nosek et al. (2015).

coups (1950-2010)[19]. In short, papers that share data get more citations, as do scholars that share their data, and datasets can get more citations than the original articles. This means that journals that require data sharing as a mandatory policy would also benefit from citations. This would start a virtuous circle of demanding that authors acknowledge the data used and thereby increasing the impact of the journal itself. Thus, it is in the self-interest of journal editors to push compulsory transparency policies in order to increase the impact of scientific publications. In

[19]Powell, and Thyne (2011) got 317 citations. See: <https://scholar.google.com/citations?user=D2RBDyAAAAAJ&hl=en> The dataset is available at: <http://www.jonathanmpowell.com/coup-detat-dataset.html>

addition, more citations for authors are naturally also good for the journals that publish their work.

After explaining why journal editors should adopt more open research policies, the next step is to present both tools and techniques.

Tools for computational reproducibility

This section presents some of the most recent tools to work transparently. Given the limitations of an essay, we briefly review the main features of each tool.

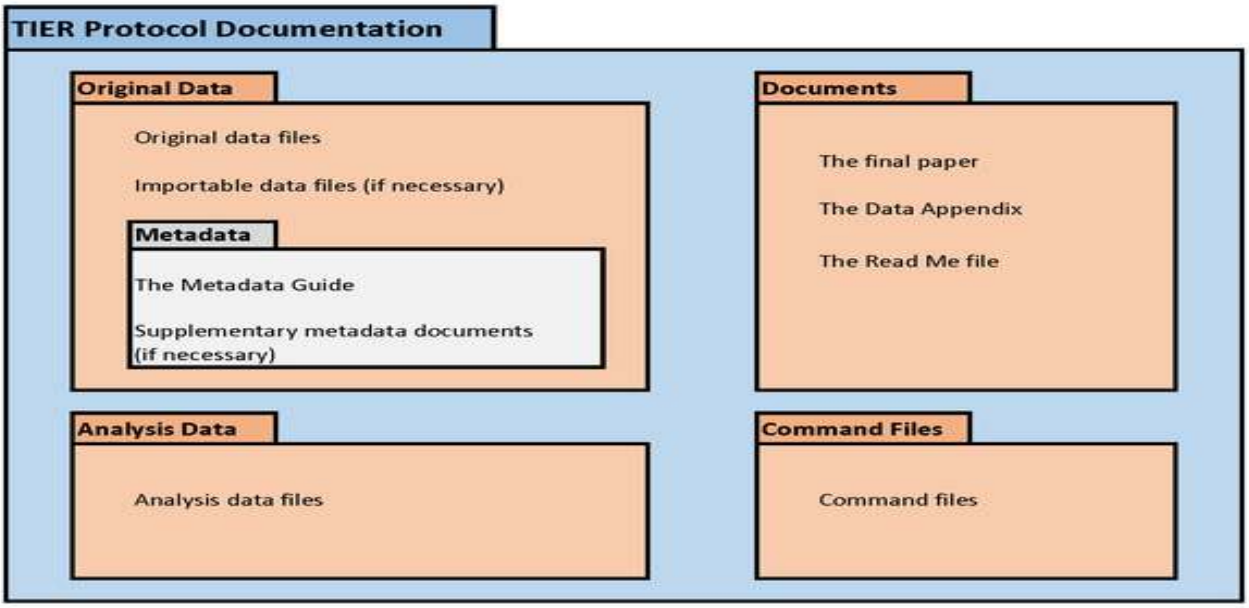
TIER Protocol

At the lowest level, you should adopt some documentation protocol to organize your files into folders. We suggest the latest version of the TIER Protocol (3.0), created by the Project Teaching Integrity in Empirical Research (TIER) and based at the Haverford College. The protocol "gives a complete description of the replication documentation that should be preserved with your study when you have finished the project" (TIER, 2017). Figure 04 shows how to organize documentation using the protocol.

In the 'original data' folder, the researcher must keep the original data used, the importable data files, and the metadata subfolder. By original data, the protocol means a copy of all the raw data used. The 'command files' folder is more computational. It should contain the code as written in the syntax of the software used in the dataset management and data analysis. Once you have a clean and processed data file, it should go into the 'analysis data' folder. Finally, the 'documents' folder: there, the researcher should keep a copy of the final work, the data appendix, and a 'readme' file. This last file should have the content of the replication documentation, changes made to importable data files, and specific instructions for study replication (TIER, 2017).

The TIER Protocol helps to organize your project. You should avoid acronyms in your files or names not directly related to your project. Following TIER recommendations, you should keep all data, codebook and script in a single folder at a single backed-up location.

Figure 04. TIER Protocol Documentation



Source: TIER (2017).

For security reasons, upload a copy at some online environment such as Dropbox, Google drive and so forth. The Open Science Framework (we will talk about them later) is an online tool that fits the protocol model perfectly. For obvious reasons, "do not spread your data over different servers, laptops and hard drives" (MARKOWETZ, 2015, p. 05)[20].

Git and Github

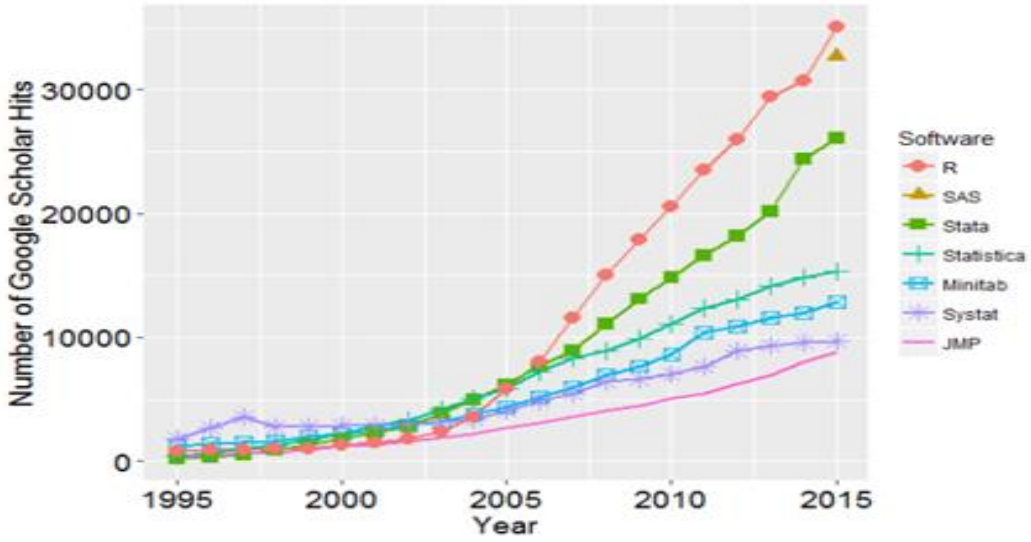
Git is a version control system (VCS) for tracking changes in computer files and coordinating work on those files among many people. Its primary purpose is software development, but it can keep track of changes in any files. Markowetz (2015) also suggests using Docker[21], which allows self-contained analysis and is easily transportable to other systems. An alternative tool to Git and Github is Mercurial, which is free and "handles projects of any size and offers an easy and intuitive interface" (GOODMAN et al., 2015, p. 05).

[20]In our teaching experience we came across a case of a master student that stored his 'analysis data' in his ex-girlfriend's laptop. Regardless of how they broke up, we strongly advise him to upload his data to an online environment as soon as possible.
[21]See: <http://www.docker.com/>.

R statistical

According to Muenchen (2012), R is starting to become the most popular programming language used for data analysis[22]. The R Project for Statistical Computing (usually named R) is an open-source and free software project widely used for data compilation, manipulation and analysis. The use of free software increases research transparency since the openness allows any observer with some basic knowledge to investigate how the author performed the analyses. In other words, using R reduces the costs of replicating work. Figure 05 illustrates its popularity over time.

Figure 05. Software hits



Source: < http://r4stats.com/articles/popularity>.

Markdown[23]

This tool, just like Latex, is another platform for document edition. However, in Markdown you can combine code, analysis and reports in a single paper. Through it, you can produce documents in many different extensions like Portable Document Format (PDF), Hyper Text Markup Language (HTML), Microsoft Word, OpenDocument Text, and Rich Text Format. Websites such as GitHub use Markdown to style their pages. The R Markdown is a commonly used

[22]See all the updates in the document at <http://r4stats.com/articles/popularity/>.
[23]See: < http://rmarkdown.rstudio.com/>.

variation of this tool that allows the manipulation of Markdown files in R's interface. Other options for integrated text and code platforms are IPython Notebook[24], ROpenSci[25], Authorea[26], and Dexy[27].

Do-Files[28]

For Stata users, the do-file is very important. When using the program, scholars can create a file containing not only all the commands used in the data analysis but also comments and instructions for reproducibility. This means that any other researcher will be able to read and reproduce your work without the need of getting in touch. Similar to the R scripts, all commands included in the Stata do-file should run with an 'execute' command. According to Richard Ball (2017), StatTag "links a Stata or SAS command file to a Word document, and then inserts tags to identify particular outputs generated by the command file and to specify where they are displayed in the Word document" (BALL, 2017)[29]. More recently, Stata 15 has enabled the conversion of dynamic Markdown documents to HTML using dyndoc[30]. Scholars can integrate text and code into a single document and share their data on the web.

Open Science Framework (OSF)[31]

According to Dessel (2017), "the Open Science Framework (OSF) is a free, open source web application that connects and supports the research workflow, enabling scientists to increase the efficiency and effectiveness of their research" (DESSEL, 2017, p. 02). On institutional grounds, OSF aims to increase openness, integrity, and reproducible research. Scholars use the OSF to collaborate, document, archive, share, and register research projects, materials, data, and pre-analysis plan. More recently, OSF launched a Pre-Registration Challenge (see next section) that is willing to grant $1,000 to scholars that register their projects before publishing the

[24]See <http://ipython.org/notebook.html>.
[25]See <https://ropensci.org/>.
[26]See <https://www.authorea.com/>.
[27]See <http://www.dexy.it/>.
[28]See: < http://www.stata.com/manuals13/u16.pdf>.
[29]See: < http://www.bitss.org/2017/01/03/stattag-new-tool-for-reproducible-research/>.
[30]See: < https://www.stata.com/new-in-stata/markdown/>.
[31]See: < https://osf.io/>.

paper. There are others management platforms, such as Taverna[32], Wings[33], and Knime[34].

Pre- analysis plan (PAP)

A pre-analysis plan is a detailed outline of the analyses that will be conducted (CHRISTENSEN and SODERBERG, 2015, p. 26). The purpose of PAP is to reduce the incidence of false positives (SIMMONS et al., 2011) and to bind our hands against data mining (CASEY et al., 2012). We suggest that a PAP should address the following features of your project: 01. sample size and sampling strategy; 02. variables measurement; 03. hypotheses; 04. full model (functional form, interactions, specification, etc); 05. methods and techniques; 06. inclusion, exclusion, transformations and corrections; 07. source of data; 08. computational software; 09. research design limitations and 10. sharing document strategy (data, codebook and scripts).

For those that wish to have a step-by-step view of a PAP, we recommend the website 'AsPredicted'[35]. There, it is necessary to answer nine questions by filling in boxes. The questions include hypothesis (including the direction of the causal effect and its strength); dependent variable; outliers and exclusions, among others. Similarly, the American Economic Association sponsors the AEA RCT Registry[36], which allows scholars to pre-register their Randomized Controlled Trials (RCT) freely.

We are not saying that these tools have to be learned all at once. Each tool requires time, and some of them display a higher learning curve, such as R Markdown. An important step is to teach these tools in both undergraduate and graduate courses, so students can learn as they acquire skills in other subjects. We firmly believe that if these tools are included in the researchers' methodological toolkit, they will have better training to answer relevant scientific questions.

[32]See <https://taverna.incubator.apache.org/>.
[33]See <http://www.wings-workflows.org/>.
[34]See <https://www.knime.com/>.
[35]See: <https://aspredicted.org/>.
[36]See: <https://www.socialscienceregistry.org/>.

Where to learn more about it?

Many sources can help scholars to learn more about reproducibility and transparency. Here, we 'emphasize' some of those with a specific focus on political science empirical research.

Data access and research transparency (DA-RT)[37]

In 2012, the American Political Science Association (APSA) Council approved the changes to the Ethics Guide to facilitate data access and increase research transparency (LUPIA and ELMAN, 2014). Article 06, for example, states that "Researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can 'be tested' or replicated" (DA-RT, 2012)[38]. In addition to 'guiding' scholars, the DA-RT initiative sponsored the election research pre-acceptance competition to foster pre-registration practices among social science scholars[39].

Berkeley initiative for transparency in the Social Sciences (BITSS)[40]

The BITSS initiative represents a significant effort to diffuse transparency in Social Science research. It provides high-quality training, prizes, research grants and educational materials to "strengthen the quality of social science research and evidence used for policy-making" (BITSS, 2017). Table 03 summarizes data on BITSS activities from 2015 to 2017.

Table 03. BITSS activities by year

Year	Catalysts	Workshops	Proposals
2015	07	04	0
2016	24	11	05
2017	70	27	36

Source: Elaborated by the authors, with data from BITSS.

[37]See: < https://www.dartstatement.org/>.
[38]See: < https://www.dartstatement.org/2012-apsa-ethics-guide-changes>.
[39]See: < https://www.erpc2016.com/>
[40]See: < http://www.bitss.org/>.

Most notably, BITSS has a catalyst program that formalizes a network of professionals to advance the teaching, practice, funding, and publishing of transparent social science research. According to the latest update, Brazil has only three official catalysts[41].

Project teaching integrity in empirical research (TIER)[42]

As we have already mentioned, the TIER Project also seeks to promote the integration of principles related to transparency and reproducibility in the research training of social scientists. It offers a development workshop at Haverford College and supports several faculty members with previous experience in incorporating the principles and methods of transparency in courses on quantitative research methods, or in the supervision of independent student research. The current version of the TIER protocol is the 3.0, and it is freely available[43].

Data repositories[44]

Harvard's 'Institute for Quantitative Social Science (IQSS) has developed the'_Dataverse Project. The principal investigator is Gary King, who collaborates with many scholars worldwide. The project is open to all disciplines and it hosts more than 371,604 files (February 2018). They offer many resources such as guidelines for replication datasets[45]. One of the most critical institutions in data sharing is the Inter-University Consortium for Political and Social Research, which aims to advance social and behavioral research. ICPSR maintains a data repository of more than 250,000 files and promotes educational activities such as the Summer Program in Quantitative Methods[46]. There are other repositories such as FigShare[47], Zenodo[48], and Dryad[49]. Some software tools may help the researchers in running

[41]See: <http://www.bitss.org/world-map-catalyst-tracker/>.
[42]See: <http://www.projecttier.org/>.
[43]Available at <http://www.projecttier.org/tier-protocol/specifications/>.
[44]See: <http://dataverse.org/>.
[45]See: <http://dataverse.org/best-practices/replication-dataset>.
[46]See: <https://www.icpsr.umich.edu/icpsrweb/sumprog/>.
[47]See <https://figshare.com/>.
[48]See <https://zenodo.org/>.
[49]See <http://datadryad.org/>.

their document repository: Invenio[50] and Eprints[51]. For qualitative scholars, the Center for Qualitative and Multi-Method Inquiry at Syracuse University sponsors the Qualitative Data Repository[52].

Journals' special issues on reproducibility

There is a growing debate on both replication and reproducibility in science. For instance, the American Psychological Association recently had an open call on the subject (submission deadline: July 01, 2018)[53]. The Political Science & Politics (PS) journal has already published two special editions on replication. The first publication was in 1995 (Volume 28) and the second in 2014: Symposium: Openness in Political Science. In 2003 (Volume 04) and 2016 (Volume 04), International Studies Perspectives published two symposia on reproducibility in international relations. In 2011, a special section of Science explored some of the challenges of reproducibility[54]. In 2016, a paper from Nature revealed, "more than 70% of researchers have tried and failed to reproduce another scientist's experiments" (BAKER, 2016, p. 452). Also in 2016, Joshua Tucker wrote in the Washington Post about the replication crises in the Social Sciences. John Oliver called attention to reproducibility in science during his Last Week Tonight[55]. In May 2017, The Annual Review of Sociology published one of the first papers on replication in the discipline[56]. The Revista de Ciencia Politica, based at the Pontificia Universidad Católica de Chile, had two articles on transparency in Volume 36, Number 03. Thad Dunning and Fernando Rosenblatt (2016) show how to make reproducible and transparent research using mixed methods. Rafael Pineiro and Fernando Rosenblatt (2016) explain how to design a pre-analysis plan for qualitative research.

[50]See: <http://invenio-software.org/>.
[51]See: <http://www.eprints.org/uk/index.php/eprints-software/>.
[52]See: <https://qdr.syr.edu/>.
[53]See: <http://www.apa.org/pubs/journals/bul/call-for-papers-replication.aspx>.
[54]See: <http://www.sciencemag.org/site/special/data-rep/>.
[55]See: <https://www.youtube.com/watch?v=0Rnq1NpHdmw>.
[56]See: <http://www.annualreviews.org/abs/doi/10.1146/annurev-soc-060116-053450>.

Coursera

Coursera is a venture-backed, for-profit, educational technology company that offers massive open online courses (MOOCs). Coursera works with universities and other organizations to make some of their classes available online, offering courses in subjects such as physics, engineering, humanities, medicine, biology, social sciences, mathematics, business, computer science, digital marketing, and data science, among others. Coursera offers data science specialization, which includes training on reproducible research[57].

Scholars' blogs

Many scholars have been pushing reproducibility forward. An important effort to advance replication is the Political Science Replication blog edited by Nicole Janz (Nottingham University)[58]. Regarding research, Janz (2016) also discusses how reproducibility can be included in current undergraduate and graduate courses. Other similar blog initiatives include Retraction Watch[59] and Simply Statistics[60].

Conclusion

This article presents seven reasons why journals should implement transparent guidelines, and what the benefits for authors are.

We have argued, in particular, that replication materials should include full disclosure of the methods used to collect and analyze data, the public availability of both raw and manipulated data, in addition to the computational scripts. Also, we have described some of the tools for computational reproducibility and reviewed different learning sources on transparency in science. We are aware that transparency does not guarantee that empirical results will be free from intentional frauds or honest mistakes. However, it makes both much less likely, since openness increases the chances of detecting such problems.

After describing the benefits of transparency, it is essential to explain what can be done to foster a more open research culture among political science scholars.

[57]See: <https://pt.coursera.org/learn/reproducible-research>.
[58]See: <https://politicalsciencereplication.wordpress.com/>.
[59]See: <http://retractionwatch.com/>.
[60]See: <http://simplystatistics.org/>.

Here we emphasize the role of the following incentives: 01. mandatory replication policies for academic journals such as the Brazilian Political Science Review; 02. specific funding for replication studies such as the SSMART grants from BITSS; 03. replication as a required component of the curriculum in undergraduate and graduate courses alike, such as Professor Adriano Codato's syllabus at Parana Federal University; 04. institutional policies to count datasets as (citable) academic outputs, such as Dataverse; 05. conferences and workshops to disseminate transparency, such as BITSS and TIER workshops; 06. development of online platforms such as the Open Science Framework, Harvard Dataverse, the ReplicationWiki (HÖFFLER, 2017) and the Political Science Replication Initiative; 07. creation of journals designed to publish successful and unsuccessful replications; and 08. publication of methodological papers related to replication, reproducibility and transparency.

Science is becoming more professional, transparent and reproducible. There is no turning back. Sooner or later, scholars who do not follow the transparency movement are likely to be left behind. Theories that are not falsifiable are destined to die, as well as opaque research practices. With this paper, we hope to foster transparency in the political science scholarly community and thus help it to survive.

Reviewed by Priscilla Kreitlon
Submitted on July 09, 2018
Accepted on November 24, 2018

References

BAGGERLY, Keith A., and COOMBES Kevin R. (2009), Deriving chemosensitivity from cell lines: forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*. Vol. 03, Nº 04, pp. 1309-1334.

BALL, Richard (2017), StatTag: new tool for reproducible research. Berkeley initiative for transparency in the Social Sciences. Available at *<https://www.bitss.org/2017/01/03/stattag-new-tool-for-reproducible-research/Ball>.* Accessed on February, 14, 2019.

BAKER, Monya (2016), 1,500 scientists lift the lid on reproducibility: survey sheds light on the 'crisis' rocking research. *Nature News*. Vol. 533, Nº 7604, pp. 452-454.

BALL, Richard and MEDEIROS, Norm (2012), Teaching integrity in empirical research: a protocol for documenting data management and analysis. *The Journal of Economic Education*. Vol. 43, Nº 02, pp. 182-189.

BELL, Mark S. and MILLER, Nicholas L. (2015), Questioning the effect of nuclear weapons on conflict. *Journal of Conflict Resolution*. Vol. 59, Nº 01, pp. 74-92.

BERG, Jeremy (2016), Editorial expression of concern. *Science*. Vol. 354, Nº 6317, pp. 1242.

BITSS, Berkeley initiative for transparency in the Social Sciences (2017), Mission and objectives. Available at <*http://www.bitss.org/about/*>. Accessed on February, 14, 2019.

CASEY, Katherine; GLENNERSTER, Rachel and, MIGUEL, Edward (2012), Reshaping institutions: evidence on aid impacts using a pre-analysis plan. *The Quarterly Journal of Economics*. Vol. 127, Nº 04, pp. 1755-1812.

CHANG, Andrew C. and LI, Phillip (2015), Is economics research replicable? Sixty published papers from thirteen journals say 'usually not'. Paper published by Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System. Available at <*http://dx.doi.org/10.17016/FEDS.2015.083*>. Accessed on February, 14, 2019.

CHRISTENSEN, Garret and SODERBERG, Courtney (2015), *Manual of best practices in transparent social science research*. Berkeley: University of California. 74 pp..

DAI, Xinyuan (2002), Political regimes and international trade: the democratic difference revisited. *American Political Science Review*. Vol. 96, Nº 01, pp. 159-165.

DA-RT, Data access and research transparency (2012), Ethics guide changes. Available at <*https://www.dartstatement.org/2012-apsa-ethics-guide-changes*>. Accessed on February, 14, 2019.

DE La O, Ana Lorena (2015), *Crafting policies to end poverty in Latin America*. New York: Cambridge University Press. 178 pp..

DE La O, Ana Lorena (2013), Do conditional cash transfers affect electoral behavior? Evidence from a randomized experiment in Mexico. *American Journal of Political Science*. Vol. 57, Nº 01, pp. 01-14.

DESSEL, Pieter Van (2017), Research integrity: the open science framework. HTTPS://OSF.IO. File. Available at <*http://www.kvab.be/sites/default/rest/blobs/1321/symposiumWI2017_OSF_VanDessel.pdf*>. Accessed on February, 14, 2019.

DEWALD, William. G.; THURSBY, Jerry G., and ANDERSON, Richard G. (1986), Replication in empirical economics: the journal of money, credit and banking project. *The American Economic Review*. Vol. 76, Nº 04, pp. 587-603.

DUNNING, Thad and ROSENBLATT, Fernando (2016), Transparency and reproducibility in multi-method research. *Revista de Ciencia Política*. Vol. 36, Nº 03, pp. 773-783.

EBERSOLE, Charles R.; AXT, Jordan R., and NOSEK, Brian A. (2016), 'Scientists' reputations are based on getting it right, not being right. *PLoS Biology*. Vol. 14, Nº 05, p.e1002460.

ELMAN, Colin and KAPISZEWSKI, Diana (2014). Data access and research transparency in the qualitative tradition. *PS: Political Science and Politics*. Vol. 47, Nº 01, pp. 43-47.

FANELLI, Daniele (2009), How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PloS ONE*. Vol. 04, Nº 05, p. e5738.

FANG, Ferric C.; STEEN, R. Grant, and CASADEVALL, Arturo (2012), Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*. Vol. 109, Nº 42, pp. 17028-17033.

FINIFTER, Bernard M. (1975), Replication and extension of social research through secondary analysis. *Social Information*. Vol. 14, Nº 02, pp. 119-153.

FRAGA, Bernard L. and HERSH, Eitan (2010), Voting costs and voter turnout in competitive elections. *Quarterly Journal of Political Science*. Vol. 05, Nº 04, pp. 339-356.

FREESE, Jeremy and PETERSON, David (2017), Replication in social science. *Annual Review of Sociology*. Vol. 43, Nº 01, pp. 147-165.

GALVÃO, Thaís Freire; SILVA, Marcus Torrentino, and GARCIA, Leila Posenato (2016), Ferramentas para melhorar a qualidade e a transparência dos relatos de pesquisa em saúde: guias de redação científica. *Epidemiologia e Serviços de Saúde*. Vol. 25, Nº 02, pp. 427-436.

GHERGHINA, Sergiu and KATSANIDOU, Alexia (2013), Data availability in political science journals. *European Political Science*. Vol. 12, pp. 333-349.

GLEDITSCH, Nils Petter and JANZ, Nicole (2016), Replication in international relations. *International Studies Perspectives.* Vol. 17, Nº 04, pp. 361-366.

GLEDITSCH, Nils Petter; METELITS, Claire, and STRAND, Håvard (2003), Posting your data: will you be scooped or will you be famous? *International Studies Perspectives*. Vol. 04, Nº 01, pp. 89-97.

GOMEZ, Brad T.; HANSFORD, Thomas G., and KRAUSE, George A. (2007), The republicans should pray for rain: weather, turnout, and voting in U.S. Presidential Elections. *Journal of Politics*. Vol. 69, Nº 03, pp. 649–663.

GOODMAN, A.; PEPE, A.; BLOCKER, A.; BORGMAN, C.; CRANMER, K.; CROSAS, M.; DI STEFANO, R.; GIL, Y.; GROTH, P.; HEDSTROM, M.; HOGG, D.; KASHYAP, V.; MAHABAL, A.; SIEMIGINOWSKA, A.; SLAVKOVIC, A. (2015), Ten simple rules for the care and feeding of scientific data. *PLOS Computational Biology*. Vol. 10, Nº 04, pp. 01-05.

GOODMAN, Steven N.; FANELLI, Danielle, and IOANNIDIS, John P. A. (2016), What does research reproducibility mean? *Science Translational Medicine*. Vol. 08, Nº 341, pp. 341ps12.

GUJARATI, Damodar (2011), *Econometrics by example*. London: Palgrave Macmillian. 400 pp..

HAMERMESH, Daniel S. (2007), Replication in economics. *Canadian Journal of Economics/Revue canadienne d'économique*. Vol. 40, Nº 03, pp. 715-733.

HERRNSON, Paul S. (1995), Replication, verification, secondary analysis, and data collection in political science. *PS: Political Science and Politics*. Vol. 28, Nº 03, pp. 452-455.

HÖFFLER, Jan H. (2017), Replication and economics journal policies. *American Economic Review*. Vol. 107, Nº 05, pp. 52-55.

IMAI, Kosuke; KING, Gary, and RIVERA, Carlos Velasco (2016), Do nonpartisan programmatic policies have partisan electoral effects? Evidence from two large scale randomized experiments. *American Journal of Political Science*. Vol. 81, Nº 02, pp. 1037a-1038a.

JANZ, Nicole (2016), Bringing the gold standard into the classroom: replication in university teaching. *International Studies Perspectives*. Vol. 17, Nº 04, pp. 392-407.

JASNY, Barbara R.; CHIN, Gilbert; CHONG, Lisa, and VIGNIERI, Sacha (2011), Again, and again, and again… . *Science*. Vol. 334, Nº 6060, p. 1225.

KEY, Ellen M. (2016), How are we doing? Data access and replication in Political Science. *PS: Political Science and Politics*. Vol. 49, Nº 02, pp. 268-272.

KING, Gary (2006), Publication, publication. *PS: Political Science and Politics*. Vol. 39, Nº 01, pp. 119-125.

KING, Gary (1995), Replication, replication. *PS: Political Science and Politics*. Vol. 28, Nº 03, pp. 444-452.

KING, Gary; KEOHANE, Robert O., and VERBA, Sidney (1994), *Designing social inquiry:* scientific inference in qualitative research. Princeton: Princeton University Press. 264 pp..

KRAWCZYK, Michal and REUBEN, Ernesto (2012), (Un) Available upon request: field experiment on researchers' willingness to share supplementary materials. *Accountability in research*. Vol. 19, Nº 03, pp. 175-186.

LEEK, Jeffrey T. and PENG, Roger D. (2015), Opinion: reproducible research can still be wrong: adopting a prevention approach. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 112, Nº 06, pp. 1645-1646.

LUCAS, Jeffrey W.; MORRELL, Kevin, and POSARD, Marek (2013), Considerations on the 'replication problem' in sociology. *The American Sociologist*. Vol. 44, Nº 02, pp. 217-232.

LUPIA, Arthur and ELMAN, Colin (2014), Openness in political science: data access and research transparency. *PS: Political Science & Politics*. Vol. 47, Nº 01, pp. 19-42.

MANSFIELD, Edward D.; MILNER, Helen V., and ROSENDORFF, B. Peter (2000), Free to trade: democracies, autocracies, and international trade. *American Political Science Review*. Vol. 94, Nº 02, pp. 305–321.

MARKOWETZ, Florian (2015), Five selfish reasons to work reproducibly. *Genome biology*. Vol. 16, Nº 01, p. 274.

MIGUEL, E.; CAMERER, C.; CASEY, K.; COHEN, J.; ESTERLING, K. M.; GERBER, A.; GLENNERSTER, R.; GREEN, D. P.; HUMPHREYS, M.; IMBENS, G.; LAITIN, D.; MADON, T.; NELSON, L.; NOSEK, B. A.; PETERSEN, M.; SEDLMAYR, R.; SIMMONS, J. P.; SIMONSOHN, U., and VAN DER LAAN, M. (2014), Promoting transparency in social science research. *Science*. Vol. 343, Nº 166, pp. 30-31.

MONTENEGRO, Renan Holanda and MESQUITA, Rafael (2017), Leaders or loners? How do the BRICS Countries and their regions vote in the UN General Assembly. *Brazilian Political Science Review*. Vol. 11, Nº 02, pp. e0005.

MONTREAL GAZETTE (2017), U. S. student offers $5,000 reward for PhD data stolen in Montreal. Montreal, November, 30, 2017. News. Available at *<http://montrealgazette.com/news/local-news/u-s-student-offers-5000-reward-for-phd-data-stolen-in-montreal>*. Accessed on February, 14, 2019.

MORAVCSIK, Andrew (2014), Transparency: the revolution in qualitative research. *PS: Political Science and Politics*. Vol. 47, Nº 01, pp. 48-53.

MUENCHEN, Robert A. (2012), The popularity of data analysis software. Available at <*URL: http://r4stats. com/popularity*>. Accessed on February, 14, 2019.

MUNAFÒ, Marcus. R.; NOSEK, Brian A.; BISHOP, Dorothy V. M.; BUTTON, Katherine S.; CHAMBERS, Christopher D.; SERT, Nathalie Perci du; SIMONSOHN, Uri; WAGENMAKERS, Eric-Jan; WARE, Jennifer J., and IOANNIDIS, John P. A. (2017), A manifesto for reproducible science. *Nature Human Behaviour*. Vol. 01, Article 21, pp. 01-09.

NATURE GEOSCIENCE, (2014), Toward Transparency. Available at < http://www.nature.com/ngeo/focus/transparency-in-science/index.html>. Accessed on February, 14, 2019.

NIH - National Institutes of Health (2018), Rigor and reproducibility in research: guidance for reviewers. Available at <https://www.med.upenn.edu/bgs-rcr-exdes/nih-ppt.pdf>. Accessed on February, 14, 2019.

NOSEK, BRIAN A.; ALTER, G.; BANKS, G. C.; BORSBOOM, D.; BOWMAN, S. D.; BRECKLER, S. J.; BUCK, S.; CHAMBERS, Christopher D.; CHIN, G.; CHRISTENSEN, G.; CONTESTABILE, M.; DAFOE, A.; EICH, E.; FREESE, Jeremy; GLENNERSTER, R.; GOROFF, D.; GREEN, D. P.; HESSE, B.; HUMPHREYS, M.; ISHIYAMA, J.; KARLAN, D.; KRAUT, A.; LUPIA, A.; MABRY, P.; MADON, T.; MALHOTRA, N.; MAYO-WILSON, E.; MCNUTT, M.; MIGUEL, E.; LEVY PALUCK, E.; SIMONSOHN, Uri; SODERBERG, C.; SPELLMAN, B. A.; TURITTO, J.; VANDENBOS, G.; VAZIRE, S.; WAGENMAKERS, Eric - Jan; WILSON, R., and YARKONI T. (2015), Promoting an open research culture. *Science*. Vol. 348, Nº 6242, pp. 1422-1425.

NSF–National Science Foundation (2015), Reproducibility, replicability, and generalization in the social, behavioral, and economic sciences. Available at <*https://www.nsf.gov/sbe/SBE_Spring_2015_AC_Meeting_Presentations/Bollen_Report_on_Replicability_SubcommitteeMay_2015.pdf*>. Accessed on February, 14, 2019.

PATIL, Prasad; PENG, Roger D., and LEEK, Jeffrey (2016), A statistical definition for reproducibility and replicability. bioRxiv. Available at doi: <*https://doi.org/10.1101/066803*>. Accessed on February, 14, 2019.

PINEIRO, Rafael and ROSENBLATT, Fernando (2016), Pre-analysis plans for qualitative research. *Revista de Ciencia Política*. Vol. 36, Nº 03, pp. 785-796.

PIWOWAR, Heather A.; DAY, Roger S., and FRIDSMA, Douglas B. (2007), Sharing detailed research data is associated with increased citation rate. *PloS ONE*. Vol. 02, Nº 03, p. e308.

POTTI, Anil; DRESSMANN, Holly K.; BILD, Andrea; RIEDEL, Richard F.; CHAN, Gina; SAYER, Robin; CRAGUN, Janiel; COTTRILL, Hope; KELLEY, Michael J.; PETERSEN, Rebecca; HARPOLE, David; MARKS, Jeffrey; BERCHUCK, Andrew; GINSBURG, Geoffrey S.; FEBBO, Phillip; LANCASTER, Johnathan, and NEVINS, Joseph R. (2006), Genomic signatures to guide the use of chemotherapeutics. *Nature Medicine*. Vol. 12, Nº 11, pp. 1294-1300.

RAUCHHAUS, Robert (2009), Evaluating the nuclear peace hypothesis: a quantitative approach. *Journal of Conflict Resolution*. Vol. 53, Nº 02, pp. 258-278.

REINHART, Carmen M. and ROGOFF, Kenneth S. (2010), Growth in a time of debt (digest summary). *American Economic Review*. Vol. 100, Nº 02, pp. 573-578.

SAVAGE, Caroline J. and VICKERS, Andrew J. (2009), Empirical study of data sharing by authors publishing in PLoS journals. *PloS ONE.* Vol. 04, Nº 09, pp. e7078.

SCHNACKENBERG, Andrew (2009), Measuring transparency: towards a greater understanding of systemic transparency and accountability. *Paper* published by Department of Organizational Behavior. Weatherhead School of Management. Case Western Reserve University. 55 pp..

SEAWRIGHT, Jason and COLLIER, David (2004), Glossary. In: *Rethinking social inquiry:* diverse tools, shared standards. Edited by BRADY, Henry E. and COLLIER, David. Lanham: Rowman and Litlefield. pp. 273-313.

SIMMONS, Joseph P.; NELSON, Leif D., and SIMONSOHN, Uri (2011), False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*. Vol. 22, Nº 11, pp. 1359-1366.

SIMONSOHN, Uri (2013), Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological science*. Vol. 24, Nº 10, pp. 1875-1888.

STOCKEMER, Daniel; KOEHLER, Sebastian, and LENZ, Tobias (2018), Data access, transparency, and replication: new insights from the political behavior literature. *PS: Political Science & Politics*. Vol. 51, Nº 04, pp. 01-05.

STRAND, Hårvard.; BINEAU, Maria Seippel; NORDKVELLE, Jonas; GLEDITSCH, Nils Petter (2016), Posting your data: from obscurity to fame? Available at <*https://osf.io/v5qm9/?action=download&version=1*>. Accessed on February, 14, 2019.

TIER (2017), Specifications. Available at <*http://www.projecttier.org/tier-protocol/specifications/*>. Accessed on February, 14, 2019.

TUCKER, Joshua (2016), Does social science have a replication crisis? The Washington Post. Washington, March 09. Available at <*https://www.washingtonpost.com/news/monkey-cage/wp/2016/03/09/does-social-science-have-a-replication-crisis/?utm_term=.f548a9d1b70f*>. Accessed on February, 14, 2019.

YOUNG, Joseph K. and JANZ, Nicole (2015), What social science can learn from the LaCour scandal. Available at <*http://www.chronicle.com/article/What-Social-Science-Can-Learn/230645*>. Accessed on February, 14, 2019.

Appendix

Table 04. Data sharing across disciplines

Author (year)	Discipline	Sample	Compliance rate (%)
Krawczyk and Reuben (2012)	Economics	200	44
Savage and Vickers (2009)	Medicine	10	10
Wicherts et al (2006)	Psychology	141	27
Wicherts, Bakker and Molenaar (2011)	Psychology	49	42,9
Wollins (1962)	Psychology	37	24
Craig and Reese (1973)	Psychology	53	37,74
Mean			30,94

Source: Elaborated by the authors.

brazilianpoliticalsciencereview

ERRATUM

In the article "Seven Reasons Why: A User's Guide to Transparency and Reproducibility", doi: 10.1590/1981-3821201900020001, at the header,

Where it reads:
"Article"

Read on:
"Review Essay"