

Data and Code Availability in Political Science Publications from 1995 to 2022

Carlisle Rainey, *Florida State University, USA*

Harley Roe, *Florida State University, USA*

Qing Wang, *Florida State University, USA*

Hao Zhou, *Florida State University, USA*

ABSTRACT

In this article, we assess the availability of reproduction archives in political science. By “reproduction archive,” we mean the data and code supporting quantitative research articles that allows others to reproduce the computations described in the published article. We collect a random sample of quantitative research articles published in political science from 1995 to 2022. We find that—even in 2022—most quantitative research articles do not point to a reproduction archive. However, practices are improving. In 2014, when the DA-RT symposium was published in *PS: Political Science and Politics*, about 12% of quantitative research articles point to the data and code. Eight years later, in 2022, that has increased to 31%. This underscores a massive shift in norms, requirements, and infrastructure. Still, only a minority of articles share the supporting data and code. In 2014, Lupia and Alter wrote, “Today, information on the data production and analytic decisions that underlie many published works in political science is unavailable.” They could write the same today; much work remains to be done.

Political science is a leader in scientific practices that ensure reproducible research (Moody, Keister, and Ramos 2022). Lupia and Elman (2014, 23) note that “openness is an indispensable element of credible research and rigorous analysis, and hence essential to both making and demonstrating scientific progress.” Despite the broad movement toward reproducibility in political science since at least King’s (1995) transformative recommendations, many quantitative research articles do not share the underlying data and code. And despite the large literature reviewing our

practices in political science (e.g., Key 2016; Stockemer, Koehler, and Lentz 2018), we do not know how the availability of reproduction files has changed since 1995. In this project, we document the current practices and describe how practices have changed since 1995. How often do publications share their data and code today? And how has this changed over the last three decades?

A growing body of literature in the social sciences emphasizes the importance of replication and reproducibility to scientific credibility (e.g., Grossman and Pedahzur 2021). In the last two decades, scholars have developed a huge body of work that discusses the replicability and reproducibility of research (for a thorough review, see National Academies of Sciences, Engineering, and Medicine 2019). This effort is gaining momentum across the social and medical sciences, and political scientists are leaders in this effort (Moody, Keister, and Ramos 2022; see also the symposium on “Openness in Political Science” in the January 2014 issue of *PS: Political Science and Politics* and the colloquium in the March 13, 2018, issue of *PNAS*).

Corresponding author: Carlisle Rainey  is an associate professor of political science at Florida State University. He can be reached at crainey@fsu.edu.

Harley Roe  is a PhD candidate in political science at Florida State University. He can be reached at hroe@fsu.edu.

Qing Wang  is a PhD candidate in political science at Florida State University. She can be reached at qwang10@fsu.edu.

Hao Zhou  is a PhD student in political science at Florida State University. He can be reached at hzz2c@fsu.edu.

© The Author(s), 2025. Published by Cambridge University Press on behalf of American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Unfortunately, this literature has sometimes offered different and even contradictory definitions of “replication” and “reproducibility.” However, an emerging consensus is that “replication” refers to researchers obtaining substantively similar results across multiple studies (i.e., using different data). Reproducibility, on the other hand, refers narrowly to computational reproducibility—that other researchers can use the same data (and perhaps even the same code) to obtain the same results. In this project, we focus narrowly on the *availability* of the data and code that support the quantitative analysis.¹ The *availability* of the data and code, in turn, facilitate *computational reproducibility*. In political science, for example, some journals require authors to share their data and computer code prior to publication.² At some journals, an editorial assistant reruns the analysis and confirms that the results match those reported in the article (Brodeur et al. 2024). For these journals that require sharing data and code, a large percentage of articles have accompanying data and code (Key 2016). However, not all journals require sharing data and code publicly. Although this policy is common for the most visible journals in political science, like the *American Journal of Political Science*, *Political Analysis*, and *International Organization*, the policies are not widespread. In this article, we document how often publications include their data and code across a wide range of political science journals from 1995 to 2022.

THE IMPORTANCE OF COMPUTATIONAL REPRODUCIBILITY

Social science is an increasingly computational science. We have seen tremendous growth in the availability and use of quantitative data (for a review, see Brady 2019). Furthermore, the availability of fast, powerful computers has led to the broad adoption of complex computational methods. For projects using large, quantitative data sets or complex computational tools, it is usually difficult (or impossible) to supply readers with the full details of the

More importantly, by sharing the data and code, researchers allow others to reproduce the results, confirm the correctness of the computation, understand undocumented decisions, and build on the research.

method. For these complex projects, the researchers must supply data and code to make the methods transparent. More importantly, by sharing the data and code, researchers allow others to reproduce the results, confirm the correctness of the computation, understand undocumented decisions, and build on the research (Barnes 2010). Buckheit and Donoho (1995, 5) make the point starkly: “An article about computational science in a scientific publication is not the scholarship itself, it is *merely advertising* of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions that generated the figures” (italics ours).

But regardless of whether the article is “merely advertising,” researchers widely share a value that researchers *ought* to allow others to verify their computations. Peng (2011, 1226) writes, “[t]he standard of reproducibility calls for the data and the computer code used to analyze the data be made available to others... [U]nder this standard, limited exploration of the data and the analysis

code is possible and may be sufficient to verify the quality of the scientific claims.”

Critically, neither the availability of a reproduction archive nor the successful reproduction of the results guarantees the *correctness* of the results. Instead, the availability of a reproduction archive makes results *verifiable* by documenting precisely how the results were created. Donoho (2010, 358) argues that “computation-based science publication is currently a doubtful enterprise because there is not enough support for identifying and rooting out sources of error in computational work.” Sharing data and code allows authors to demonstrate the correctness of the results (and introduces a powerful incentive for correctness) and allows others to verify their claims.

EXISTING EFFORTS TO ASSESS AVAILABILITY

There are two main existing efforts to assess availability of reproduction archives in political science. Key (2016) examines the availability of reproduction archives for articles published in 2013 and 2014 in *American Political Science Review*, *American Journal of Political Science*, *British Journal of Political Science*, *International Organization*, *Journal of Politics*, and *Political Analysis*. She thoroughly searched for publicly posted material but did not email the authors to request the reproduction archive. Across these top journals in 2013 and 2014, she finds that about 66% of articles publicly share the underlying data and code. However, she finds considerable heterogeneity across journals. Among the journals that *required* sharing a reproduction archive, 93% of articles supply the underlying data and code. But among journals that did not require sharing, the number drops to about 43%. However, Key (2016) focuses on six of the most visible journals in political science, and some of these have among the most aggressive policies requiring public reproduction archives. Gherghina and Katsanidou (2013) looked at the *policies* employed by a large collection of 120 political science

journals beyond the highly visible journals in Key’s (2016) study. Of these 120 journals, only 18 had data availability policies posted on their website.

Stockemer, Koehler, and Lentz (2018) take a different approach. They identify all articles published in three political behavior subfield journals: *Electoral Studies*, *Party Politics*, and *Journal of Elections, Public Opinion, and Parties*. None of these journals required sharing data and code at the time of their study. For each article, they carefully searched for replication data, and if their search was unsuccessful, they contacted the authors (up to four times) to request the data. Despite a laborious search, Stockemer, Koehler, and Lentz were able to obtain the data and code for only about 57% of the articles in their study. The authors identified 145 articles and found posted reproduction archives for 13 (9%) and obtained archives via email for 69 (48%). Despite their thoroughness, these authors are unable to obtain reproduction archives for 43% of the articles published in these reputable journals.

This existing work leaves an important question remaining. Key (2016) finds that sharing is common for articles in the top journals, but only when those journals require sharing. For a set of subfield journals focusing on political behavior (that do not require sharing), Stockemer, Koehler, and Lentz (2018) find that only about 9% of articles publicly share their data and code. This raises an important question: For a broad collection of political science journals, how common is publicly sharing the data and code that supports an article? And how has that changed over time?

To address these questions, we explore the availability of reproduction archives across a much wider time frame (1995 to 2022) and a much wider set of journals (all English-language journals in the Social Science Citation Index's "Political Science" and "International Relations" categories). Our findings are much less optimistic than Key's (2016) results but perhaps more optimistic than those of Stockemer, Koehler, and Lentz's (2018). To preview, we find that only about 31% of articles published in 2022 point readers to reproduction archives. That rate is steadily shrinking, but incrementally.

DATA

To explore the availability of reproduction archives across a wider time frame and set of journals, we take a random sample of quantitative research articles published in political science journals from 1995 to 2022 and code whether the publication points readers to publicly available data and code. We discuss the details of this procedure below.

To obtain the sample of quantitative research articles, we proceed in three steps:

1. First, we generate a list of political science journals—a surprisingly challenging step. We rely on the Social Science Citation Index, using their "Political Science" and "International Relations" categories. As a practical matter, we remove all 18 journals in languages other than English, leaving 224 journals. We use Crossref's API to collect all 560,514 digital object

The changes shown in Figure 1 represent a massive shift in norms, requirements, and infrastructure. And the improvement is not accidental but due to a decades-long, deliberate effort by leaders in the field.

identifiers (DOIs) from these 224 journals from 1995 to 2022. From this list, we take a simple random sample of 5,000 DOIs and an additional over-sample of 1,000 DOIs from 2022.

2. A team of four coders assess each of the 6,000 DOIs. We code whether the DOI belongs to a research article with underlying quantitative data and computer code that the authors could potentially share—what we call a "quantitative research article." Many DOIs do not meet this criterion; we discard those. We are left with a random sample of 1,413 quantitative research articles.
3. We evaluate whether each of the 1,413 quantitative research articles points readers to the underlying data and code. We classified the articles into the following categories based on whether and how they shared their underlying data and code: not mentioned, available upon request, description of where

to find (but hard to find), description of where to find (and easy to find), linked (but no longer available), or linked (and still available).³ Two caveats are important. First, if the article does not point the reader to the data and code, we code that article as "not mentioned."⁴ This standard aligns with discipline norms and King's (1995, 446) standard that researchers should share data and code with a professional archive (e.g., ICPSR at the time) and that "it should be made publicly available and reference to it made in the original publication (usually in the first footnote)." Second, we did not evaluate the quality of the files; we only evaluated whether the publication points to the data and/or code. If the publication points to the data and code, we consider the archive "available" regardless of whether the data and code are working or complete.

RESULTS

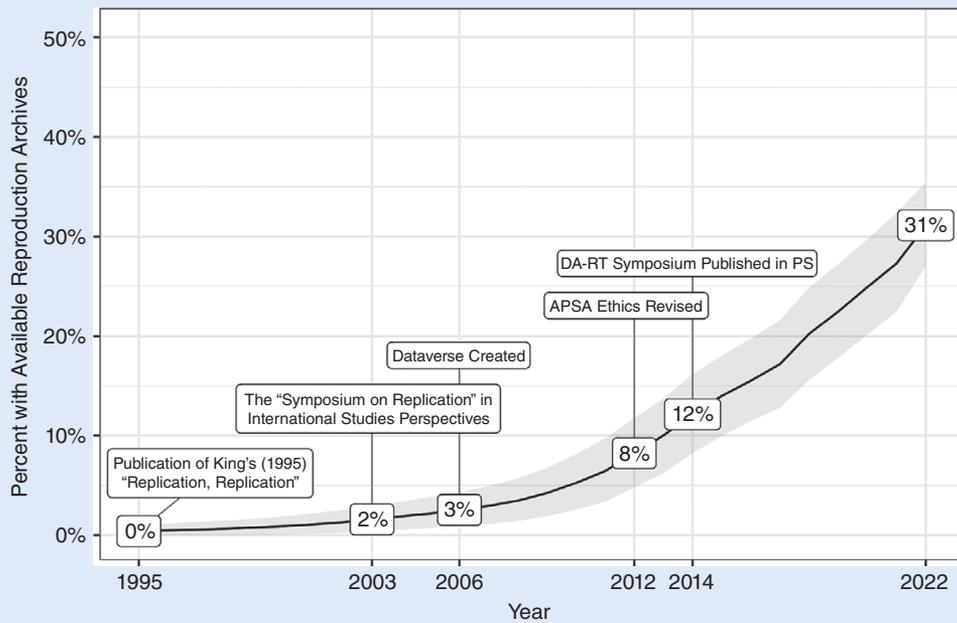
To show how practices are changing over time, we start by collapsing the cases into two categories: shared or not. If the publication (a) provides a working link to the reproduction archive or (b) describes where to find the reproduction archive and it was easy to find, then we consider that data and code "available." Otherwise, we consider it "unavailable." To increase the precision of the estimates, we fit a regression model that estimates a monotonic increase each year in the rate of availability.⁵

Figure 1 shows the percentage of articles that share data and code over time. Starting with the publication of King's (1995) "Replication, Replication," almost no articles shared their data and code. We code 122 articles from 1995 to 2002, and none contained working links to data and code or described how to find it—the first article in our sample to share data and code occurred in 2003. Figure 1 shows that sharing remains relatively rare (less than about 5%) until about 2008, when availability starts to become more common. For context, *International Studies Perspectives* published a symposium in 2003 in which the editors of four prominent international relations journals (*Journal of Peace Research*, *International Studies Quarterly*, *Inter-*

national Interactions, and *Journal of Conflict Resolution*) urge other editors to join them in requiring that that authors "must make their data available" (Bueno de Mesquita et al. 2003). The Dataverse Project began three years later in 2006 (King 2007). From about 2008 to the present, sharing has become more common, although perhaps not as rapidly as one might think. When APSA released its professional ethics standard in 2012 (APSA 2012), about 8% of articles shared their data and code. That number was about 12% in 2014 when Lupia and Alter (2014, 57) wrote, "[t]oday, information on the data production and analytic decisions that underlie many published works in political science is unavailable." A decade later, that number has almost tripled to 31%. Still, 69% of articles do not share their data and code, but we have made *incremental* progress. The changes shown in Figure 1 represent a massive shift in norms,

Figure 1

Percentage of Quantitative Articles with Reproduction Archives, 1995–2022



This figure shows the estimated percentage of quantitative research articles published in political science journals from 1995 to 2022 that supply reproduction archives.

requirements, and infrastructure. And the improvement is not accidental but due to a decades-long, deliberate effort by leaders in the field (e.g., King 1995; Bueno de Mesquita et al. 2003; King 2007; Lupia and Alter 2014).

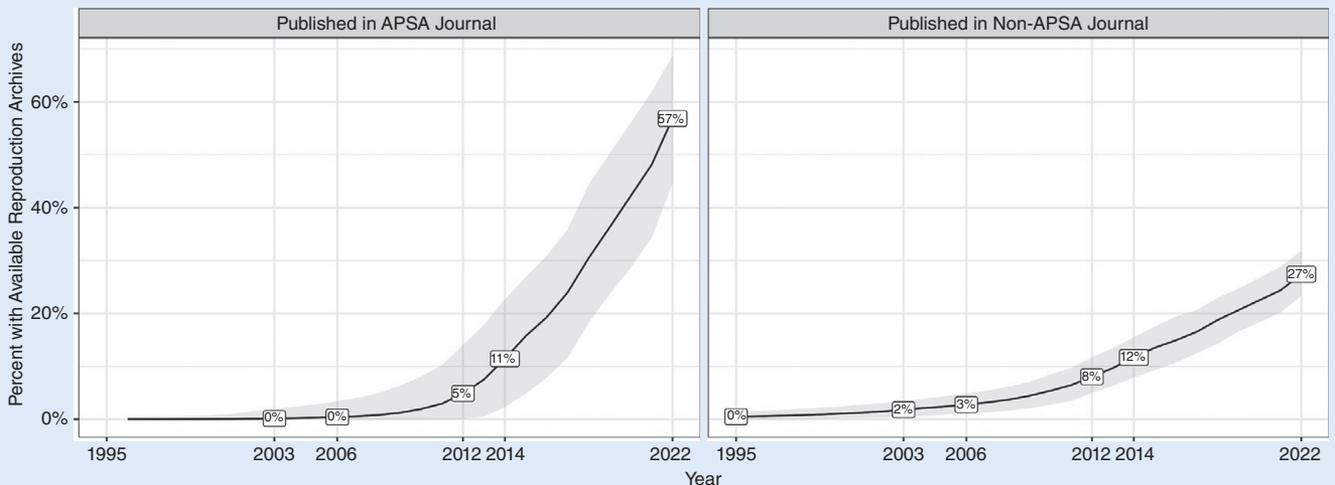
Next, we break the articles into categories to look for heterogeneity in practices. First, we look at APSA journals. These journals include the organization’s journals (*American Political Science Review*, *Perspectives on Politics*, *PS: Political Science and Politics*) and the section journals (e.g., *Political Behavior*, *State Politics and Policy Quarterly*). We might expect large differences

between the APSA and non-APSA journals. After all, APSA journals are among the most visible journals in the discipline, and the APSA has a formal code of ethics that requires sharing data and/or code to reproduce the results (absent special circumstances).

As one might expect, Figure 2 shows that the percentage of articles published in 2022 with reproduction archives available is much higher for APSA journals (57%) than for non-APSA journals (27%). However, barely a majority (57%) of articles published in our discipline’s core journals share their data and code with

Figure 2

Results for APSA and non-APSA Journals



This figure shows the estimated percentage of quantitative research articles published in APSA journals and non-APSA political science journals from 1995 to 2022 that supply reproduction archives.

the community; 43% do not. This underscores our point that improvement in sharing practice has been *incremental* and that *more work remains to be done*, even in the journals at the center of our discipline.

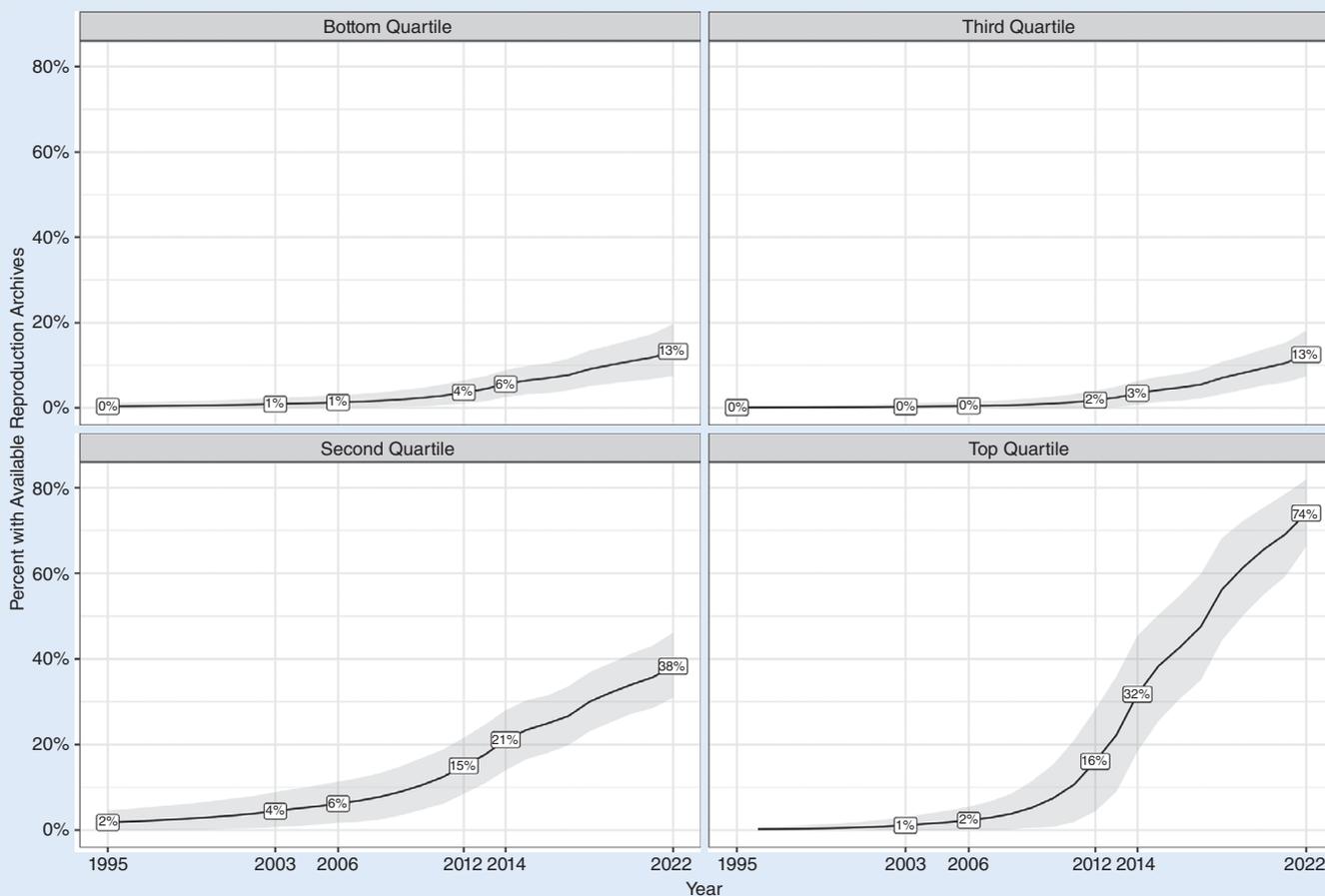
Next, we break the articles into four categories based on journal rankings. We use the SCImago Journal Rank score from 2022 to give each journal a rating. Then we place the 1,413 articles into one of four equally sized bins based on the SCImago Journal Rank score of the journal they were published in. As examples, articles published in *American Political Science Review*, *International Organization*, or *Political Science Research and Methods* are placed in the top quartile, those in *Journal of Peace Research* or *Electoral Studies* in the second, those in *Conflict Management and Peace Science* or *Legislative Studies Quarterly* in the third, and those in *International Interactions* or *State Politics and Policy Quarterly* in the bottom category. One might expect that articles published in the highest ranked journals to have higher rates of sharing. Indeed, that is what we find.

Figure 3 shows that about 74% of articles published in 2022 in journals ranked in the top quartile supply a reproduction archive. However, sharing rates drop dramatically outside the top quartile—which includes many high-profile journals (10 of

the 14 APSA journals in our data fall outside the top quartile). About 38% of articles in the second quartile supply reproduction archives. The percentage in the third and bottom quartiles drops to 13% each.⁶

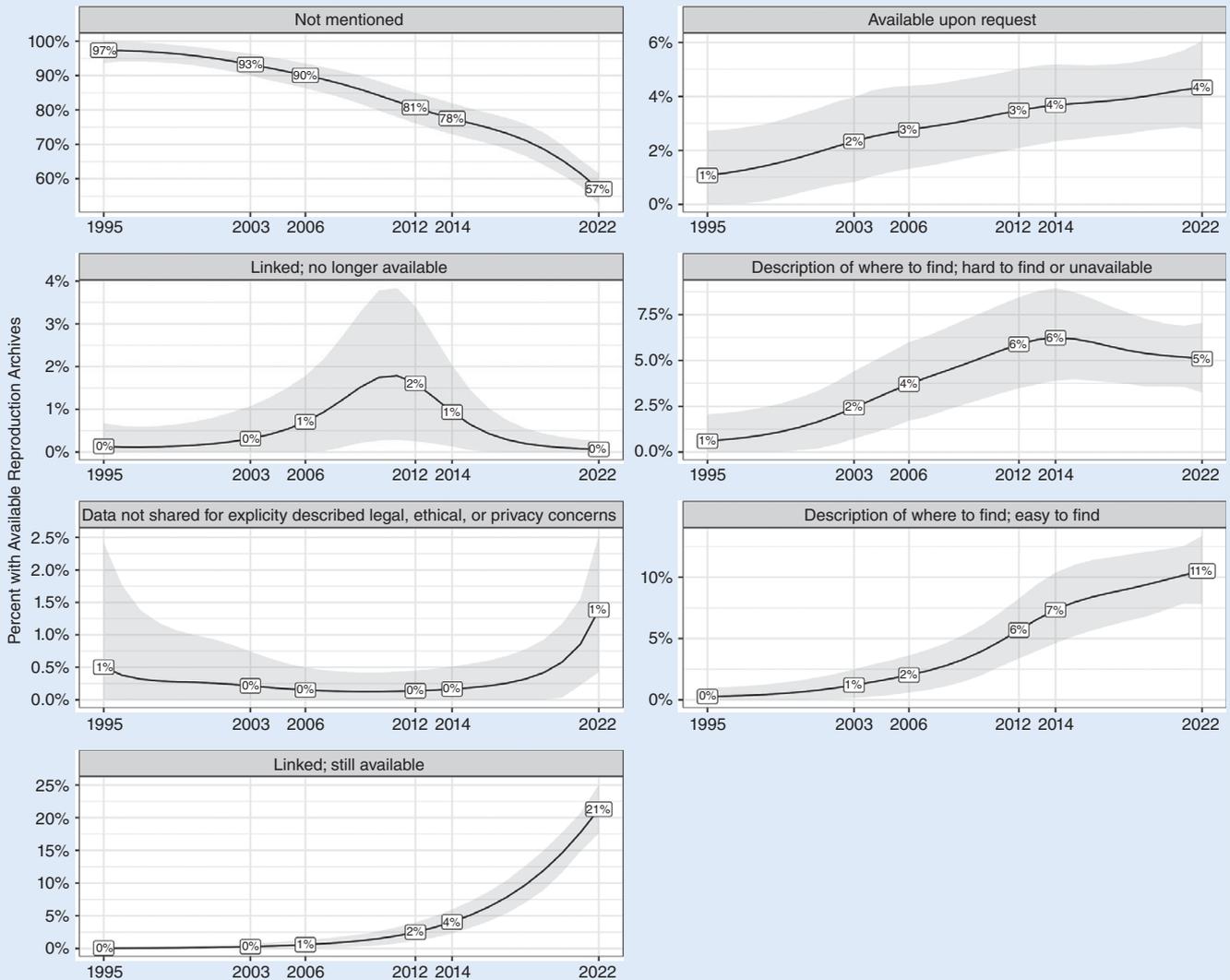
Last, Figure 4 shows the results for the broad range of categories that we coded. Most of the movement happens between the two important categories of “not mentioned” and “linked, still available.” The number of articles that do not mention the data and code shrinks from about 97% in 1995 to 57% in 2022. We have seen progress (but incremental). The number of articles with working links to their reproduction archives has grown from 0% in 1995 and 4% in 2014 to 21% in 2022. About 11% of articles published in 2022 offer a description of where to find the archive, and we were able to easily locate it. This number continues to grow; we encourage authors to instead use a permanent archive (e.g., OSF or Dataverse) and supply a persistent DOI for the archive with their published article. The number of articles that have data “available upon request” is growing but low—about 4% in 2022. The percentage with a link that no longer works is surprisingly low, about 2% for articles published around 2010, which suggests that professional archives like Dataverse have been remarkably successful. About 5% of recent articles offer a description of where to find the

Figure 3
Results by Journal Ranking Quartile



This figure shows the estimated percentage of quantitative research articles published in variously ranked political science journals from 1995 to 2022 that supply reproduction archives.

Figure 4
Results for Disaggregated Categories



This figure shows the estimated percentage of quantitative articles published in political science journals from 1995 to 2022 that falls into our seven categories.

reproduction archive (e.g., “available on the author’s website”), but we were not able to easily locate the archive (e.g., the author updated their website URL). The number of authors that opt out of sharing data and/or code for explicitly stated legal or ethical

revisions to APSA ethics (discussed in the 2014 symposium in *PS*) formally affirm this value. But the question remains: How are we doing? Do researchers consistently share their data and code? Is sharing increasing?

In 2014, when the DA-RT symposium was published, about 12% of quantitative research articles point to the data and code. Eight years later, in 2022, that has increased to 31%, underscoring a massive shift in norms, requirements, and infrastructure.

concerns remains negligible—less than about 1%—throughout the period.

CONCLUSION

King (1995) implores political scientists to share their data and code so that others can verify and build on their results. The 2012

To address these questions, we take a random sample of 6,000 articles published in political science journals, identify 1,413 for which sharing data and code is appropriate, and code whether the published article points to a publicly available reproduction archive. Unfortunately, we find that—even in 2022—the answer is “usually not.” However, practices are incrementally improving.

In 2014, when the DA-RT symposium was published, about 12% of quantitative research articles point to the data and code. Eight years later, in 2022, that has increased to 31%, underscoring a massive shift in norms, requirements, and infrastructure. Still, only a minority of articles share the supporting data and code. And the results seem robust. Since we completed our data collection, Scoggins and Robertson (2024) published similar data using an alternative approach and obtain largely similar results.

We suspect that these changing practices are driven by changing journal requirements. Key (2016) shows that journal requirements are a primary driver in sharing rates. Although verifying the replication code can be costly and require additional organization, Key (2016) notes that editors can easily require authors to include a persistent link to the reproduction archive in the final manuscript—this is the core policy recommended in the *International Studies Perspectives* symposium more than two decades ago (Bueno de Mesquita et al. 2003). Although this minimal policy does not guarantee that the results are reproducible, it “allows other interested scholars to verify and use the data and code and provides an opportunity for students to learn through replication” (Key 2016, 271). In this article, we show that for too many articles published today, others cannot verify the results and students cannot use the data and code to learn through replication. We encourage editors to (continue to) insist on it.

DATA AVAILABILITY STATEMENT

The editors have granted an exception to the data policy for this manuscript. In this case, replication code and data are available to reproduce its figures and tables, but there are substantively small differences between the replication and the printed results. This exception was granted because the authors affirmed that these differences are attributable to randomness in the sampling procedure that generates draws from Bayesian posterior distributions that do not change the conclusions of the manuscript.

CONFLICTS OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research. ■

NOTES

1. Although we focus on the data and code for *quantitative* research projects, a large body of literature also focuses on the importance of sharing *qualitative* data as well. See Kapiszewski and Karcher (2020, 2021) for reviews.
2. Of course, sometimes authors cannot legally or ethically make their data available to others.
3. In rare instances, the authors stated that they would release the data after an embargo period or described a specific reason they could not share their data (e.g., privacy concerns). We noted these cases.
4. For example, if the author posted the reproduction archive to Dataverse but does not mention this in the article, then we code this as “not mentioned.” But if the author includes a statement such as “data and code are available on the author’s website,” then we made a good-faith effort to locate the reproduction files.
5. We fit these monotonic regression models using Stan with the *brms* package in R (Bürkner and Charpentier 2020). Estimating a smooth change over time that is not necessarily monotonic does not alter the substantive conclusions.
6. We suspect that variation in journal *policies* explains most of the variation across the journal ranking categories. To assess this suspicion, we collected data on the current policy stated on journal website. However, given the long timeline between submission and publication, we could not connect our data on journal policies to specific publications. Although the correlation was consistent with our suspicion, we omit this analysis because it can easily be misleading. To minimize the

potential for misinterpretation, we present the data on journal policies in Rainey and Roe (2024). See Brodeur et al. (2024) for similar data.

REFERENCES

- APSA (American Political Science Association). 2012. *Proposed Revised Text to the Guide to Professional Ethics in Political Science, Section III, Principles of Professional Conduct, Section A, Principles for Individual Researchers*. <https://www.apsanet.org/portals/54/Files/Proposed%20Changes%20to%20Ethnic%20Guide.pdf>.
- Barnes, Nick. 2010. “Publish Your Computer Code: It Is Good Enough.” *Nature* 467 (7317): 753–53.
- Brady, Henry E. 2019. “The Challenge of Big Data and Data Science.” *Annual Review of Political Science* 22 (1): 297–323.
- Buckheit, Jonathan B., and David L. Donoho. 1995. “Wavelab and Reproducible Research.” In *Wavelets and Statistics*, ed. Anestis Antoniadis, 55–81. New York: Springer-Verlag.
- Bürkner, Paul-Christian, and Emmanuel Charpentier. 2020. “Modelling Monotonic Effects of Ordinal Predictors in Bayesian Regression Models.” *British Journal of Mathematical and Statistical Psychology* 73 (3): 420–51. <https://doi.org/10.1111/bmsp.12195>
- Bueno de Mesquita, Bruce, Nils Petter Gleditsch, Patrick James, Gary King, Claire Metelits, James Lee Ray, Bruce Russett, Håvard Strand, and Brandon Valeriano. 2003. “Symposium on Replication in International Studies Research.” *International Studies Perspectives* 4 (1): 72–107. <https://doi.org/10.1111/1528-3577.04105>
- Brodeur, Abel, Kevin Esterling, Jörg Ankel-Peters, Natália S. Bueno, Scott Desposato, Anna Dreber, Federica Genovese, et al. 2024. “Promoting Reproducibility and Replicability in Political Science.” *Research and Politics* 11 (1). <https://doi.org/10.1177/20531680241233439>
- Donoho, David L. 2010. “An Invitation to Reproducible Computational Research.” *Biostatistics* 11 (3): 385–88.
- Gherghina, Sergiu, and Alexia Katsanidou. 2013. “Data Availability in Political Science Journals.” *European Political Science* 12 (3): 333–49.
- Grossman, Jonathan, and Ami Pedahzur. 2021. “Can We Do Better? Replication and Online Appendices in Political Science.” *Perspectives on Politics* 19 (3): 906–11.
- Kapiszewski, Diana, and Sebastian Karcher. 2020. “Making Research Data Accessible.” In *The Production of Knowledge: Enhancing Progress in Social Science*, ed. Colin Elman, John Gerring, and James Mahoney, 197–220. Cambridge: Cambridge University Press.
- Kapiszewski, Diana, and Sebastian Karcher. 2021. “Transparency in Practice in Qualitative Research.” *PS: Political Science and Politics* 54 (2): 285–91.
- Key, Ellen M. 2016. “How Are We Doing? Data Access and Replication in Political Science.” *PS: Political Science and Politics* 49 (2): 268–72.
- King, Gary. 1995. “Replication, Replication.” *PS: Political Science and Politics* 28 (3): 444–52.
- King, Gary. 2007. “An Introduction to the Dataverse Network as an Infrastructure for Data Sharing.” *Sociological Methods and Research* 36 (2): 173–99.
- Lupia, Arthur, and Colin Elman. 2014. “Openness in Political Science: Data Access and Research Transparency.” *PS: Political Science and Politics* 47 (1): 19–42.
- Lupia, Arthur, and George Alter. 2014. “Data Access and Research Transparency in the Quantitative Tradition.” *PS: Political Science and Politics* 47 (1): 54–59.
- Moody, James W., Lisa A. Keister, and Maria C. Ramos. 2022. “Reproducibility in the Social Sciences.” *Annual Review of Sociology* 48 (1): 65–85.
- National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press.
- Peng, Roger D. 2011. “Reproducible Research in Computational Science.” *Science* 334 (6060): 1226–27.
- Rainey, Carlisle, and Harley Roe. 2024. “The Data Availability Policies of Political Science Journals.” SocArXiv, working paper dfzya. <https://doi.org/10.31235/osf.io/dfzya>
- Scoggins, Bermond, and Matthew P. Robertson. 2024. “Measuring Transparency in the Social Sciences: Political Science and International Relations.” *Royal Society Open Science* 11: article 240313. <https://doi.org/10.1098/rsos.240313>
- Stockemer, Daniel, Sebastian Koehler, and Tobias Lentz. 2018. “Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature.” *PS: Political Science and Politics* 51 (4): 799–803. <https://doi.org/10.1017/S1049096518000926>