Data access and research transparency (DA-RT) is a growing concern for the discipline. Technological advances have greatly reduced the cost of sharing data, enabling full replication archives consisting of data and code to be shared on individual websites, as well as journal archives and institutional data repositories. But how do we ensure that scholars take advantage of these resources to share their replication archives? Moreover, are the costs of research transparency borne by individuals or by journals? This article assesses the impact of journal replication policies on data availability and finds that articles published in journals with mandatory provision policies are 24 times more likely to have replication materials available than articles those with no requirements.

The controversy surrounding LaCour and Green (2014) highlights the importance of replication and verification. The inability of researchers to replicate the central findings (Broockman, Kalla, and Aronow 2015) and the subsequent retraction of the article by Science editors caused a scandal in the field and beyond—similar to the aftermath of the discovery of Reinhart and Rogoff’s (2010) spreadsheet error in economics (Herndon, Ash, and Pollin 2013). These alleged errors, and others like them, were identified using publicly available replication archives. The public availability of these archives, however, is largely due to efforts made by journals to increase research transparency.

Data access and research transparency (DA-RT) is a growing concern for the discipline. Technological advances have greatly reduced the cost of sharing data, enabling full replication archives consisting of data and code to be shared on individual websites, as well as journal archives and institutional data repositories. But how do we ensure that scholars take advantage of these resources to share their replication archives? Moreover, are the costs of research transparency being borne by individuals or by journals? Expanding on the work of Gherghina and Katsanidou (2013), I move from the journal-level to the article-level to assess the impact of journal replication policies on data availability. I conclude with suggestions for increasing research transparency.

**EXISTING EFFORTS**

The goal of publishing replication archives is not simply internal verification or the correction of sloppy scholarship. Rather, replication also allows for extension through the collection of new data and the application of different methods (Fowler 1995; King 2006). Although scholars have an incentive to ensure that their data are available and up to date as a way to increase exposure and citation counts (Gleditsch, Metelits, and Strand 2003), it is difficult to achieve compliance on a voluntary basis (Anderson et al. 2005; King 1995). Recognizing that relying on scholars to self-police is suboptimal, journals have recently created or revised their replication policies to advance social rather than individual responsibility. In other words, the burden is shifting to editors to ensure the availability of replication archives for work published in their journal (Gherghina and Katsanidou 2013; Ishiyama 2014).

Part of this shift is due to journals committing to the DA-RT statement developed by the APSA council (APSA 2014). Based on the “principle that sharing data and information fuels a culture of openness that promotes effective knowledge transfer” (Lupia and Elman 2014, 20), editors of DA-RT journals require data to be uploaded to a journal repository at the time of publication. There are many benefits of these repositories, including durable, central archives that do not require individuals to be responsible for maintenance. Older, more prestigious, and general-interest journals are more likely to have replication policies than those with lower-impact factors or more specific audiences (Gherghina and Katsanidou 2013). This is a self-reinforcing process because more readily available data increases citation counts, thereby boosting the impact factor of a journal.

Journal policies that require replication may affect material availability beyond an author’s natural tendency to publish replication archives. Ensuring that replication standards are met, however, strains scarce journal resources. If scholars are already maintaining complete replication archives on their own, there is no need for editors to police their authors. If replication policies are not fully enforced, the effort expended for partial enforcement may be wasted (Dafoe 2014).
That is, simpler policies that require less effort from journal editors appear to be as effective as more resource-intensive verification policies.
Web searches for an author(s) were used when articles did not include any mention of replication archives or contained broken links, or when replication materials were not otherwise located. Those searches led to websites that contain replication materials for 77 articles, 68 of which contained both data and code. In other words, 26.4% of the total replication materials found were discovered through virtual digging; the need to search so thoroughly makes the replication processes more difficult than necessary.

More than 40% of links to websites included in articles were broken, which indicates that authors believe their replication materials are available when in reality they are not. This is particularly problematic for personal websites, especially when authors have changed institutions since publication of their articles. Moreover, these figures are based on recently published articles. As articles age, the likelihood of a “dead” or broken link increases. If scholars forgo dataverses and other durable archives, they must take extra care in maintaining their own websites.

As noted previously, some type of replication materials—data, code, or both—are available for 327 articles in the sample and full archives for 292. The strongest predictor of availability is whether a journal has a policy mandating that data and/or code be made publicly available at the time of publication (table 4). By requiring replication archives, it is 24 times more likely that any materials will be provided and 17 times more likely that a full replication package will be published. This echoes the claim of Gherghina and Katsanidou (2013, 337) that “[t]he most important element of a data availability policy is the extent it binds the authors.” Likewise, journals with less stringent policies (i.e., JOP, BJPS, and IO before 2014) are more likely to have articles with replication archives than those that do not have replication requirements.

**Table 2**

Replication Material Availability by Policy

<table>
<thead>
<tr>
<th>Replication Policy</th>
<th>Data and Code</th>
<th>Only Data</th>
<th>Only Code</th>
<th>Not Available</th>
<th>Total Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verification</td>
<td>82.7%</td>
<td>3.4%</td>
<td>3.4%</td>
<td>10.3%</td>
<td>29</td>
</tr>
<tr>
<td>Mandatory Provision</td>
<td>88.4%</td>
<td>2.3%</td>
<td>2.3%</td>
<td>70%</td>
<td>172</td>
</tr>
<tr>
<td>Expected Provision</td>
<td>43.1%</td>
<td>7.6%</td>
<td>2.2%</td>
<td>47.3%</td>
<td>68</td>
</tr>
<tr>
<td>Not Required</td>
<td>27.9%</td>
<td>4.4%</td>
<td>0%</td>
<td>67.6%</td>
<td>68</td>
</tr>
</tbody>
</table>

**Table 3**

Replication Material Availability by Source

<table>
<thead>
<tr>
<th>Source</th>
<th>Data and Code</th>
<th>Only Data</th>
<th>Only Code</th>
<th>Dead Links</th>
<th>Total Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal</td>
<td>75.5%</td>
<td>4.9%</td>
<td>2.5%</td>
<td>17.2%</td>
<td>212</td>
</tr>
<tr>
<td>Repository</td>
<td>68.9%</td>
<td>4.9%</td>
<td>1.6%</td>
<td>24.6%</td>
<td>61</td>
</tr>
<tr>
<td>Website Linked in Article</td>
<td>50.0%</td>
<td>9.5%</td>
<td>0%</td>
<td>40.5%</td>
<td>42</td>
</tr>
<tr>
<td>Searched for Website</td>
<td>81.0%</td>
<td>7.1%</td>
<td>3.6%</td>
<td>8.3%</td>
<td>84</td>
</tr>
<tr>
<td>Other External Website</td>
<td>18.2%</td>
<td>45.5%</td>
<td>0%</td>
<td>36.4%</td>
<td>11</td>
</tr>
</tbody>
</table>

More than 40% of links to websites included in articles were broken, which indicates that authors believe their replication materials are available when in reality they are not.

The age of an article, measured as a count of the number of quarters since publication, does not have a significant effect on the likelihood of data sharing. It does not appear that authors find time to provide replication materials in the months following publication; neither has the discipline’s recent focus on replication influenced the probability that newer articles will be published with replication materials. Rather, the degree of data access far more depends on a publishing outlet’s policies.
DISCUSSION AND RECOMMENDATIONS

As with any collective action, diffusion of responsibility leads to shirking; the same is true for DA-RT. More than 33% of articles in the sample did not have publicly available replication materials; an additional 7% provided only some of the information needed for replication. To ensure greater cooperation, external enforcement mechanisms are necessary. The previous analyses confirm that the extent to which replication archives are provided is largely a function of journals requiring research transparency.

Although these replication policies are effective in increasing compliance, shifting the burden of research transparency to journals is costly. Whereas verification of the analyses presented in an article before publication is the “gold standard,” it is unreasonable—and likely unnecessary—for all journals to implement such rigorous policies. Many journals lack editorial assistants, leaving the certification of results to editors. Considering the volume of submissions, verification of analyses is not feasible except at well-staffed journals. In addition to concerns about efficient allocation of editorial time and effort, editors and staff may not have access to every program or add-on used in an analysis. Furthermore, simply because results can be verified using the data and code provided does not assure situations in which the data contain serious errors or are somehow falsified.

Rather than verifying analyses before publication, journals should model their replication policies after journals such as PA, which requires that specific replication materials be uploaded to the journal’s dataverse and cited in an article’s references. This allows other interested scholars to verify and use the data and code and provides an opportunity for students to learn through replication (Janz 2015). It also relieves journals from the burden of duplicating results while still requiring that materials be made publicly available.

Even with mandatory provision policies, the compliance rate falls short of 100%. What are we to make of the approximately 20% of more substantive pieces that fail to fully comply with journal policies? The lack of availability may simply be an oversight on the part of authors or it may stem from a lack of appreciation for the importance of replication to the field as a whole. APSA, this journal, and others in the discipline stress the benefits of data access and replication, but the message has not reached everyone. Rather than devoting resources to the verification of results, journals can improve availability by certifying that authors have complied with replication policies before publication.

It is important to note that replication files for this analysis were downloaded but not opened or run and therefore may not be complete. By coding articles based on the availability rather than the integrity of the replication package, this article assesses only whether a minimum standard is being met. Journals should establish specific guidelines about the contents of a full replication archive (Altman and King 2007; Eubank 2014). The APSA section responsible for its journal also should maintain the journal’s dataverse, alleviating work for overburdened editors. Last, there is a need for archives to be associated with articles through persistent identifiers rather than web links (Ishiyama 2014). In summary, data access cannot be the sole responsibility of individual researchers. Journals must take a more active role in building a culture of data sharing and ensuring research transparency.

NOTES

1. See McCullough and McKitrick (2009) for a list of studies that failed to replicate results across a variety of fields.
2. Although arguably more difficult, there have been calls for qualitative studies to subscribe to the same standards at quantitative work (Elman and Kapiszewski 2014; Golden 1995).
3. APSA requires citations to be in the first footnote, whereas PA includes data citations in the reference list.
4. An additional 17 articles contained broken links to personal websites.
5. Political Science Research and Methods is an example of a less-established journal that is able to require verification of analyses before publication.
6. Replication materials for this article are available at http://dx.doi.org/10.7910/DVN/5IJAMC.

REFERENCES


Table 4

Logistic Regression of Factors Influencing Replication Material Availability

<table>
<thead>
<tr>
<th></th>
<th>Any Materials</th>
<th>Full Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (Std. Error)</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>Mandatory Provision</td>
<td>3.25* (0.30)</td>
<td>25.8</td>
</tr>
<tr>
<td>Expected Provision</td>
<td>0.86* (0.18)</td>
<td>2.3</td>
</tr>
<tr>
<td>Quarters Since Publication</td>
<td>−0.02 (0.08)</td>
<td>1.0</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.83 (0.34)</td>
<td>0.5</td>
</tr>
<tr>
<td>N = 494</td>
<td>71.05%</td>
<td>71.46%</td>
</tr>
</tbody>
</table>

Notes: *p < 0.05. Standard errors clustered by journal.
The Profession: Date Access and Replication in Political Science


