Report for 2020 by the AEA Data Editor

By LARS VILHUBER*

The American Economic Association (AEA) Data Editor's mission is to "design and oversee the AEA journals' strategy for archiving and curating research data and promoting reproducible research" (Duflo and Hoynes, 2018). The 2018 Report by the Data Editor (Vilhuber, 2019) articulates how to implement that mission. Since July 2019, we have conducted comprehensive pre-publication reproducibility checks for all regular AEA journals, developed guidance for authors, and worked with peers at societies and groups in economics and elsewhere.

Based on the experience from the first full year of reproducibility verification, we implemented several improvements in mid-2020. We provided improved guidance to authors depositing materials, clarified the data and code availability policy (DCAP), expanded and clarified additional policies on third-party reproducibility checks, on post-publication updates to replication packages, and on required replication materials for field and lab experiments (Section I). We expanded pre-publication reproducibility checks to all AEA journals, implemented some basic checks on replication packages for Papers and Proceedings, and worked on expanding verification capacity and usability of the data repository (Section II). In the course of 2020, we reached out to numerous data creators and providers — both authors who have created unique data resources, and academic and commercial data providers that often provide the data for economic research — and have discussed with these data providers access to data for reproducibility checks, mechanisms to request publication approval, and generally

informed them of the need for reproducibility, provenance tracing, and transparency in economic research. We continue to coordinate with other journals, societies, and registries on these topics (Section III).

I. Updating the AEA Data and Code Availability Policy

After having introduced stronger principles of provenance and transparency into the AEA's data and code availability policy in 2019 (American Economic Association, 2020), we monitored compliance and understanding as evidenced by the replication package submissions we received. It became quickly clear that stronger guidance, and greater clarity were needed to assist authors in complying with the DCAP. Authors struggled with how best to document their code and data, with the process of deposing data and code, and the ability to provide clear data provenance, including data citations.

To address these issues and improve compliance upon submission, we took a two-pronged approach: We clarified the policy, and provided improved instructions and guidance on how to comply with the policy. We released the revised version in September 2020. The main content remains unchanged, but we simplified the main policy, separated out the policy as applied to papers conducting (field and lab) experiments, and expanded the policy to encompass more clearly any primary data collection. We also introduced supplementary policies that lay out how and when we expect reproducibility checks by third-parties to be conducted (also see our interactions with third-party verifiers, and, as a logical

* Cornell University, lars.vilhuber@cornell.edu.

consequence of more transparent data and code deposits, how to revise data and code deposits when post-publication changes become necessary. These policies are reprinted in American Economic Association (2021) and accessible online at https://www.aeaweb.org/journals/data.

We also clarified the data and code availability policy as it applies to papers submitted to these Proceedings. All empirical and computational articles in the proceedings need to make code and data available in the same AEA Data and Code Repository as articles in the journals. However, other than verifying compliance with the policy (materials were deposited) and conducting basic checks (material can be read, metadata is in order), we do not conduct the same pre-publication reproducibility checks as we do for regular articles (see Section I.E for statistics on compliance).

But policies only outline what should occur, not how authors can achieve compliance. In order to assist authors in complying with the policy, thus improving turnaround times, we developed more detailed step-by-step guidance at aeadataeditor.github.io/aea-de-guidance/step-by-step.html. We also revised the information that authors need to provide when their manuscript is conditionally accepted (or, in some cases, during the review process). We have eliminated the duplicative requirement to provide README files during the manuscript submission process — they remain a required element in the data and code deposit process — replacing it with a simpler form that requires authors to attest to compliance of their manuscript with the data citation requirement, provide details on where the data and code were deposited, and has them alert the Data Editor to any constraints on data access[1] At the same time, the form provides hyperlinks to the guidance mentioned above. A simplified version of the form

(reduced to a few checkboxes) is incorporated into the initial manuscript submission workflow. The primary goal of these forms and checkboxes is to alert authors to the need to comply with the policy early on, and hopefully leading to better compliance by the time that manuscripts are conditionally accepted. We have already observed some improvement in the few months since these improvements were fully rolled out.

Furthermore, in collaboration with data editors at other journals in economics , management, and statistics (see Section III), we have released a template README (Vilhuber et al., 2020) which helps authors compile all the information required for complete documentation of their data and code deposit.[2] With the publication of this report in the Proceedings, we will start verifying that authors use the README to document their replication package.

A. Improving Findability of Replication Materials at the AEA

Since the AEA began using a data and code repository hosted by the Inter-university Consortium for Political and Social Research (ICPSR), all deposits have JEL codes, and authors can add funder information and key metadata describing the data that is available within the deposit. Many authors avail themselves of this option. These attributes aid in making each deposit findable through scholarly search engines such as Web of Science[3] and Google Scholar.[4]

Deposits made prior to July 2019 and migrated to the AEA Data and Code Repository in October and December 2019 do not have this rich metadata, since it must be entered interactively by authors. To improve the metadata of older deposits, we ran a campaign in summer and fall of 2020, thanks to the support provided

[1] The form is available at aeaweb.org/journals/forms/data-code-availability.

[2] The README is available at social-science-data-editors.github.io/template_README/.

[3] https://clarivate.com/webofsciencegroup/solutions/web-of-science/

[4] https://scholar.google.com/

by a research team at the University of Michigan and ICPSR. This campaign targeted several hundred papers, providing authors with a tool to augment the metadata associated with their deposits. A report on the success of the campaign is forthcoming, and the metadata provided by the authors is being incorporated as this report was being compiled. A second campaign will complete the work in Spring or Summer of 2021.

Deposits receive their own Digital Object Identifier (DOI), and can be cited in line with the Data Citation Principles (Altman and Crosas, 2013; Data Citation Synthesis Group, 2014). Authors are required to explicitly cite their data supplement if it contains original data; an entry in the references is automatically added in all cases. Conversely, the article that a supplement is associated with is clearly identified on the supplement's landing page. Future enhancements to the platform will be able to display other articles that also cite a supplement.

We continue to verify that authors cite the datasets they have used and accessed, as required by the DCAP, the AEA's citation requirements, and in line with the Data Citation Principles. Data citations increase findability of data, allow data providers to receive proper credit, and align the Association with broader principles in the academic publishing world. The AEA's Sample References[5] provide a style reference, and additional guidance for non-standard data sources, such as confidential or proprietary data, has been developed in collaboration with other journals and guidance by librarians.[6]

B. *Third-party repositories*

The DCAP allows for code and data to be deposited at other trusted repositories, as long as all other elements of the DCAP are complied with. In fact, authors are discouraged from duplicating deposits they have made elsewhere in the AEA Data and Code Repository. This is intended to allow authors to create replication packages prior to submitting at the AEA's journals, or any other journal, as a component of a reproducible workflow and possibly in compliance with funder data management policies. We have not observed many authors availing themselves of this possibility, with probably less than 10 deposits made in such a way. Authors wishing to deposit replication packages early in the research lifecycle are encouraged to consult the Social Science Data Editors website,[7] where links to trusted repositories are provided.

C. *Post-publication modifications*

At the time of publication, the article is linked with one (or more) archived supplements, constituting the version of record. What happens when, despite pre-publication scrutiny, an error in the code emerges? The DCAP provides for a clear policy, identifying which modifications constitute a minor edit to the version of record, and which modifications lead to a higher version number, without modification of the existing version of record. In particular, any change that potentially changes a computational result or adds (untested) code will lead to a new version of the deposit being created, without changing or removing the version of record, even if the modifications fixes an error. However, the presence of replication packages that are newer than the version of record is signalled to readers via a banner, and is recorded in the metadata.

We have had few such post-publication updates. A handful of authors have requested that their historical migrated packages be updated, typically with an improved README. In a few cases, we had published incomplete replication packages when access to the code was not fea-

---

[5]https://www.aeaweb.org/journals/policies/sample-references

[6]https://social-science-data-editors.github.io/guidance/addtl-data-citation-guidance.html

[7]https://social-science-data-editors.github.io/

sible due to COVID-19 related office closures. When authors regained access only after publication of the manuscript, the replication package was updated in a transparent fashion, compliant with our policy. Moving forward, the ability to transparently update replication packages to accommodate improvements or fixes may also be leveraged to reduce the time between conditional acceptance and publication.

D. Intellectual Property and Licenses

For new deposits at the AEA Data and Code Repository, authors retain copyright. The default license for all repositories based at openICPSR is the Creative Commons Attribution (CC-BY) (Creative Commons, 2017), but authors can choose their own license, though such licenses will be vetted by the Data Editor for compliance with the DCAP.

We see few authors deviate from the standard license. When deviating, authors tend to choose a different, open source license (Open Source Initiative, 2018) for software and code. In a very small number of cases, we have worked with authors to publish data under more restrictive licenses, due mostly to ethical concerns, while ensuring that replication remains possible.

On the other hand, we regularly inquire with authors about their ability to redistribute data they obtained from data providers. Some (commercial) data providers require a subscription, and disallow any redistribution of more than a dozen data points. Some (academic) data providers, while providing data for free to most researchers, do not allow for redistribution of some or all of their data. Rights to redistribute, or prohibition of redistribution, is mostly captured in "licenses" or "data use agreements", sometimes in "terms of use." We have, on numerous occasions, had to ask authors to remove data extracts that infringed on copyrights or terms of use. In one case, we had to de-accession a replication package (remove the ability to download) as a consequence of a formal notice

from a data provider of an infringing dataset. We continue to assist authors in remaining compliant with data use agreements and copyright law, to the extent possible, but authors should be aware of their potential liability in the cases of infringements.

E. Compliance

We note that authors can be compliant with the policy without providing a copy of data used, as long as the reason for the inability to provide the data is acceptable, correct, and documented as part of the replication materials.

Compliance with the policy (Table 1) has been excellent. In some cases, we have requested data that was not initially provided, when such data could be legally and ethically provided; by the second round of assessments, compliance was generally achieved (see also our discussion of outreach to data providers). In a very small number of cases, authors having conducted research in restricted-access data environments did not provide code. In several cases, relying on personal experience or by reaching out to the managers of the restricted-access environment, we were able to inform the authors of their ability to provide code, in compliance with any security requirements of the restricted-access data environment. In several cases, authors were unable to access the restricted-access environment due to COVID-19 lockdowns, which we accommodated in various ways; however, in all cases, code was ultimately provided or is expected. For a very small number of articles, compliance was achieved after the manuscript had been published; 5such documents were pending as of 2019-12-01. We are aware of only one case where the authors were unable to provide code because their employer asserted copyright and did not give permission to post code, making the deposit non-compliant.

Table 1—: Compliance

	Compliant	Incomplete	Non-compliant
AER and journals	259	1	0
Papers and Proceedings	80	4	1

Note: Unit of observation are manuscripts assessed between 2019-12-01 and 2020-11-30. "Incomplete" means the data deposit has not been completed, though the article has been published. "Non-compliant" means that code or other required materials have not been provided.

II. Infrastructure for Verification of Reproducibility

The Data Editor manages the infrastructure needed to access data and code, conduct reproducibility checks, and archive and preserve replication packages. The first two points are provided by the replication team at Cornell University, the latter primarily by the AEA Data and Code Repository provided by openICPSR at the University of Michigan, with additional support from the AEA's in-house IT staff.

A. Pre-publication verification of computational reproducibility

THE PROCESS

Pre-publicaton verification is conducted by the Data Editor's team at Cornell University. Requests for assessment of reproducibility are received, assigned to a team member, who then assesses data availability and compliance with requirements. When some data is available, a full or limited reproducibility check is conducted. If we cannot obtain access to the data or computational resources in a timely fashion, we may reach out to third-parties who can, and request a reproducibility check from them. Once all computations have been completed, a process that can take anywhere from a few minutes to several weeks, a report is compiled, reviewed and approved by the Data Editor, and submitted back to journal editors, who handle most communications with the authors. The report will have one of four possible recommendations (see Table). A "conditional acceptance" implies that a revision will need to be resubmitted to the Data Editor to address any identified shortcomings. An "acceptance" means that no further changes are necessary, and both the manuscript (after copy-editing) and the replication package can be scheduled for publications.[8] However, to streamline processing, we may also recommend an "acceptance with modifications requested." In such a case, the remaining modifications are minor, and can be handled during copy-editing (for instance, a small number of tables need minimal changes) and prior to publication of the replication package (for instance, a fixable error in a program, or a clarification in the README, not affecting any important tables or figures). We have increasingly made use of this feature. While we monitor that authors comply with the request for modifications, no further computational assessment is made. A recommendation of "revise and resubmit" is recorded when we receive a request prior to a conditional acceptance, i.e., during the "R&R" phase. When we have serious concerns, we will reach out directly to the responsible editor, and discuss solutions with the authors. In the past year, two manuscripts required conversations with the editors and the authors, and both were resolved to the satisfaction

---

[8]Manuscript and replication package are generally published at the same time, though at the request of either editors or authors, the replication package can be published at any time after acceptance.

of all involved.

ASSESSMENTS MADE

Between 2019-12-01 and 2020-11-30, the AEA Data Editor team received 832 requests, for 509 manuscripts.[9] Requests typically are channeled to the team by the AEA's journal submission and review system, but others were initiated by authors or editors directly, often while preparing the replication materials. Of these, 715 reports (402 manuscripts) were submitted back to editors,[10] and 307 were completed up to the point of publication of the data deposit, including any post-acceptance modifications. Table 3 shows the distribution of the last recommendation on record for manuscripts as of 2020-11-30. Table 2 breaks these numbers down by journal, showing the number of requests received ("rcvd") and reports completed ("cplt") in the left panel. The right panel shows the number of manuscripts for which one or more requests were received ("rcvd") and reports completed ("cplt"). The columns "external" / "ext." identify cases where we reached out to external replicators, which we discuss later. Finally, the last column identifies manuscripts for which the entire process has been completed, and which are "pending" publication.

FREQUENT ISSUES

We provide a heuristic overview of notable issues encountered when attempting to reproduce results. We note that none of the 509 manuscripts we assessed had fundamental flaws — all problems identified so far have been fixable (and fixed).

---

[9]This includes only requests submitted between those dates, and does not take into account in-progress requests on 2019-12-01.

[10]The balance are either in progress or are not coded in the adminstrative system as having been submitted to ScholarOne, such as replication packages for Papers and Proceedings.

Computational issues: We still encounter mostly preventable issues of computational instability (see McCullough and Vinod (1999) for one of the earlier observations in this vein). In most cases, authors fail to set a random seed, though this is a requirement of the policy. In some cases, differences in software versions can generate numerical differences despite having fixed a pseudo-random seed generator, for instance optimization routines in Matlab[11] and Julia v1.5's faster random number generator.[12] We have also run into issues when Fortran compilations require specific versions of libraries, when versions of Dynare are not backward-compatible, or when simple data import (Stata) or export (Matlab) functionality does not work the same way on different operating systems. In the past year, we did not encounter use of newer technologies such as "containers," and almost never the use of "manifest"-like files in R, Python, and Julia, which alleviate some of these issues.

Incomplete instructions and manual manipulation: While mechanisms such as `make` files and similar systems have been in use in software development for decades, they are very rarely used in replication packages in economics, even when it would seem a natural fit (Fortran compilations). We have encountered very manual (and tedious) instructions on how to assemble a compilable project for Fortran code, or how to generate a map using GIS software, despite the availability of reproducible tools in all cases. We encountered instructions to repeatedly change small parameters in multiple places in code prior to recompilation, or to manually move datasets between two directories, when the use of parameter files, function calls, or relative directory paths are basic mechanisms of programming. These inefficient mechanisms, while at least initially tolerated, are not just wasteful

---

[11]mathworks.com/matlabcentral/answers/422349-fmincon-in-new-matlab-version-gives-different-results

[12]docs.julialang.org/en/v1/NEWS/

Table 2—: Stats

Journal	Requests			Manuscripts			
	(rcvd)	(cplt)	(external)	(rcvd)	(cplt)	(ext.)	(pend.)
AEJ:Economic Policy	122	115	13	70	65	8	51
AEJ:Macro	131	126	3	67	64	2	46
AEJ:Micro	81	78	4	45	44	3	33
AER	182	176	12	108	103	10	80
AER:Insights	56	55	0	31	31	0	26
JEL	33	31	1	18	17	1	16
JEP	28	27	0	19	18	0	13

Notes: Data for requests received by the AEA Data Editor between 2019-12-01 and 2020-11-30. AEA P+P are excluded from this table. See text for details.

Table 3—: Recommendations

	Frequency
Accept	177
Accept - with Changes	172
Conditional Accept	41
Conditional Accept - Manuscript Ready	2
Revise and Resubmit	10

of replicator time, they are an inefficient use of the author's time as well, and can (and has) lead to errors being introduced into the manuscript's tables and figures.

Incomplete data provenance and data availability: Most articles still provide imprecise or incorrect information regarding access to data that is not provided. In some cases, authors fail to provide data that should be provided, in other cases, authors inadvertently provide data for which they do not have redistribution rights.

DELAYS

A recurring concern expressed by editors and staff members are delays in publication, due to the verification process. Table 4 shows that the median number of times a manuscript is reviewed by the AEA Data Editor is 2. Not enough manuscripts are given a rating of "acceptable"

upon first pass. This leads to substantial time before the manuscript can proceed to copy-editing.

PROCESS IMPROVEMENTS

In order to increase the initial acceptance rate, we set out in Summer 2020 to review the entire process that authors face. We made several changes aimed at (a) providing authors with the information as early as possible, when it is still easy to include reproducible practices in projects at relatively low cost and (b) providing authors with better information, to reduce frictions and uncertainty.

As noted earlier, we introduced informational forms upon submission, and a new short form, requested upon conditional acceptance, collecting salient information about the replication package. Each of these forms has links to updated and detailed guidance on how to pre-

Table 4—: Assessment rounds for completed manuscripts

Rounds	AEJ Micro	AER	AEJ Applied	AEJ Macro	AEJ Policy	AER Insights
1	15	26	22	4	20	7
2	21	47	20	35	31	16
3	5	9	3	8	7	2

Notes: Data for requests received by the AEA Data Editor between 2019-12-01 and 2020-11-30. AEA P+P are excluded from this table. See text for details.

pare and submit replication packages. We also provide links to the reproducibility checks that our team uses, so authors know in advance what checks will be applied to their replication package. These changes were rolled out in September 2020. In addition, a new template "README" (Vilhuber et al., 2020) was introduced in December 2020 as a joint effort with other data editors, and will hopefully also reduce uncertainty about content of such a README, and increase the completeness of the information upon first submission. First impressions, both at the AEA's journals as well as at the Canadian Journal of Economics, suggest that the README is achieving that goal.

B. Migrating Historical Supplements

We migrated the bulk of historical data (and code) supplements in 2019, from ZIP files stored on AEA servers to the AEA Data and Code Repository. A few hundred replication packages are still awaiting migration, pending reconciliation of duplicates and large-file limitations. Upon demand from readers, we have also sometimes reached out to authors to request materials from articles which were published prior to the original publication of a data availability policy.

Thanks to support from a research team at the University of Michigan, we were able to enlist the aid of authors of migrated replication package to improve the attached metadata. As part of the project, we sent 8,538 email to 3,005 authors of 1,460 studies. Out of the 1,000 authors (35%) who responded, 911 (30%) provided additional metadata for 522 studies. An article is expected at the conclusion of the experiment.

C. Improving the AEA Data and Code Repository

We continually assess the technical performance of the AEA Data and Code Repository. One of the pain points was a default quota limit of 2GB. While the median package size is only 24.58 MB, a significant number of authors (15 percent of deposits) had larger packages, requiring manual adjustments, and resulting in delays and occasionally confusion. However, only 1 percent of deposits were larger than 20GB. In collaboration with ICPSR, the default quota has been increased to 30GB, streamlining the process. Authors with larger packages, or packages with more than 1,000 files, should contact the AEA Data Editor.

A new workflow has also been developed, which streamlines the process for journal managers on the back-end, and prevents errors that users frequently made. It lays the groundwork for further enhancements in the future.

III. Working with Other Providers of Scientific Infrastructure to Improve Support for Documenting Provenance and Replicability

An important responsibility of the AEA Data Editor is to interact with other providers of scientific infrastructure. This involves other pub-

lishers and journals, archives such as ICPSR, providers of restricted or proprietary data, the AEA RCT Registry, metadata harvesters, and third-party verification services.

A.   Highlighting Data Resources

While reviewing replication packages, we occasionally identify data, prepared in the course of the research, which could be of greater value if separately archived and professionally curated. This may be because of significant value-added, or because the data would have been difficult to access, if not for the author's data collection or data preparation efforts. Examples include digitized maps of India made amenable for GIS software (Harari, 2020) or historical Economic Census data for the United States (Ganapati, 2020). These data are then processed by professional data librarians at ICPSR, and ultimately made available to the broader research community via the regular (non-AEA) ICPSR repository, while providing proper attribution to the author. The collaboration of the authors is voluntary, and we appreciate their efforts. Authors who believe that their data is unique and of value should contact the AEA Data Editor early enough in the research process, even when acceptance at the AEA journals may be uncertain.

B.   Data Providers

We regularly meet and communicate with academic and commercial data providers, on behalf of specific authors or because we have identified a data provider as a frequently used resource. Discussion topics include making data citations easier, clarifying licenses, requesting blanket or streamlined data redistribution authorizations, or suggesting improved data curation practices to avoid repeatedly copying data from uncurated websites to the curated AEA Data and Code Repository. As an example, after an email exchange with the leadership of the Economic Policy Uncertainty (EPU) website (Baker, Bloom and Davis, 2016), the website changed its license to a liberal Creative Commons license, which allows broad re-use and facilitates reproducibility.

In the course of the past year, we have talked about some or all of these topics with IPUMS, EPU, the UNCTAD Division on International Trade and Commodities (TRAINS data), the World Bank, Standard and Poor's, Bloomberg, Tick Data, Haver Analytics, the Federal Reserve of St. Louis data librarians,[13] the World Management Survey, Ben Zipperer at the Economic Policy Institute, Jim Poterba and Dan Feenberg at the NBER, Lucia Foster (U.S. Census Bureau), Barbara Downs (U.S. Census Bureau and Federal Statistical Research Data Centers), the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB), the German Research Data Center of the SOEP, Research Data Center of the German Pension Insurance, the UK Data Archive, the Ministério da economia do Brasil, and the French *Centre d'accès sécurisé de données* (CASD).

C.   Economics Journals

We continue to coordinate with other data editors conducting similar activities at other journals. An informal mailing list managed by the AEA Data Editor is used occasionally to interact with others.[14] Mailing list members who wish to be more actively involved can participate in the development of the website of the Social Science Data Editors.[15] The website contains guidance on data citations and data availability statements, best practices for coding and data preparation, and links to various tools useful to replicators. The recently released template

---

[13]I am indebted to Adrienne Brennecke at the St. Louis Fed, who facilitated many of these conversations.

[14]Journal editors are encouraged to join the mailing list by contacting the AEA Data Editor.

[15]socialsciencedataeditors.github.io

README is one example of collaboration, being co-authored with Miklos Kóren (Review of Economic Studies), Joan Llull (Economic Journal), Peter Morrow and Marie Connolly (Canadian Journal of Economics), and having benefited from contributions from Chris Paciorek (JASA), Ben Greiner (Management Science), Kevin Lang (JOLE), Harald Uhlig (JPE). The Data Editor has also engaged with a newer group of data repository leaders at Data-PASS organized as the Journal Editors Discussion Interface (JEDI[16]).

D. *Third-party verification services*

We continue to rely on and have discussions with third-party verification services. As noted earlier, 36 reports were provided by external replicators or replication services for 27 manuscripts (see Table 2 for statistics by journal, Appendix B for a list of third-party replicators). Of those, several were provided by institutions that already are organized as a reproducibility service: R-squared at Cornell University, cascad in France. We appreciate the willingness of all third-party replicators to provide us with their time and effort in reproducing papers. In particular, cascad generously provided us with 21 reports. The AEA Data Editor has had preliminary discussions with several institutions about the interest and possibilities of formalizing such services. Issues of cost, frequency, speed, and at what point such services would be involved in the research lifecycle remain unresolved.

IV. Working with the Economics Community to Enhance and Broaden Education on Replicable Science

We have already noted the outreach to other journals and repositories above. Educating the current and future generations of economists is one of the longer-term goals for improving re-

producibility of research. The AEA Data Editor has presented at Project TIER (Teaching Integrity in Empirical Research), and at various regional economics conferences that connect to communities of junior and later-career economists. The Data Editor participated in the AEA Committee on the Status of Women in the Economics Profession (CSWEP) "Fireside chats[17]" on January 19, 2021 with the topic "Demystifying Replication Requirements and Processes: Best Practices Viewed by AEA Data Editor." Various other recordings of presentations by the AEA Data Editor are available on Youtube, linked from the Data Editor's website.[18] Together with BITSS (UC Berkeley), a framework and platform to support teaching reproducibility is being developed, complementing other existing services such as the ReplicationWiki. Presentations at the CTREE conference are planned.

V. Data and Code Availability Statement

All publicly available data and code used to generate figures and tables in this article can be found at LINK TBD. Some detailed data from the editorial system, used for Table 4, are considered confidential and cannot be made available in a way that preserves the privacy of the editorial process at this time.

REFERENCES

Altman, Micah, and Mercè Crosas. 2013. "The Evolution of Data Citation: From Principles to Implementation." *IASSIST Quarterly*, 62–70.

American Economic Association. 2020. "Data and Code Availability Policy." *AEA Papers and Proceedings*, 110: 776–78. https://doi.org/10.1257/pandp.110.776.

[16]https://dpjedi.org/

[17]https://www.aeaweb.org/about-aea/committees/cswep/programs/resources/webinars

[18]https://aeadataeditor.github.io/talks/

American Economic Association. 2021. "Data and Code Availability Policy." *AEA Papers and Proceedings*, 111: xxx. https://doi.org/10.1257/pandp.111.xxx.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. "Measuring Economic Policy Uncertainty*." *The Quarterly Journal of Economics*, 131(4): 1593–1636. https://doi.org/10.1093/qje/qjw024.

Creative Commons. 2017. "About The Licenses." https://web.archive.org/web/20181208161819/https://creativecommons.org/licenses/ (accessed 2018-12-08).

Data Citation Synthesis Group. 2014. "Joint Declaration of Data Citation Principles." , ed. Maryann Martone. Force11. https://doi.org/10.25490/A97F-EGYK.

Duflo, Esther, and Hilary Hoynes. 2018. "Report of the Search Committee to Appoint a Data Editor for the AEA." *AEA Papers and Proceedings*, 108: 745. https://doi.org/10.1257/pandp.108.745.

Ganapati, Sharat. 2020. "Growing Oligopolies, Prices, Output, and Productivity." *American Economic Review*, forthcoming. https://www.aeaweb.org/articles?id=10.1257/mic.20190029.

Harari, Mariaflavia. 2020. "Cities in Bad Shape: Urban Geometry in India." *American Economic Review*, 110(8): 2377–2421. https://doi.org/10.1257/aer.20171673.

McCullough, B. D., and H. D. Vinod. 1999. "The Numerical Reliability of Econometric Software." 37(2): 633–665. https://doi.org/10.1257/jel.37.2.633.

Open Source Initiative. 2018. "Open Source Licenses by Category." https://web.archive.org/web/20181208161736/https://opensource.org/licenses/category (accessed 2018-12-08).

Vilhuber, Lars. 2019. "Report by the AEA Data Editor." *AEA Papers and Proceedings*, 109: 718–29. https://doi.org/10.1257/pandp.109.718.

Vilhuber, Lars, Marie Connolly, Miklós Koren, Joan Llull, and Peter Morrow. 2020. "A template README for social science replication packages." https://doi.org/10.5281/zenodo.4319999. The README is under a CC-BY-NC license. Usage by commercial entities is allowed, reselling it is not.

Replication team at Cornell University

The following students have provided excellent assistance in reproducing the results from the 509 articles processed by the Replication Lab: Ololade Omotoba, Craig Schulman, Daniella Pena, Jeong Hyun Lee, Jiayin Song, Jill Nicole Crosby, Joshua Gil Passell, Julia R Zimmerman, Justin Sun, Lauren Stubbs, Kate Hofer, Kirubeal Wondimu, Liam Cushen, Louis Liu, Luis Lopez Cabrera, Lydia Reiner, Manas Kumar Gogula, Mary-Jo Ajiduah, Matt H Wang, Melanie Chen, Nehedin Juarez, Franklin Omullo, Peter Rafael Sanchez, Rubal Mistry, Naomi Li, Ryan Ali, Seong Hwan Kim, Siyang Yu, Steve Yeh, Syon Verma, Tarangana Thapa, Valerie Setiawan, Weilun Shi, William Silverman, Xiangru Li, Ximei Shen, Yanyun Chen, Yuan-Hsuan Lin, Zebang Xu, Zechariah Karsana, Zhaojiahong Zhu.

Graduate students David Wasser, Meredith Welch, and Hyuk Harry Son (all Cornell University) and Research Aide Michael Darisse (Cornell University) have been invaluable assistants in training and coordinating the work as well as developing the methods and procedures which we have made public. Harry Son contributed programming to this report.

Third-party replicators

We are grateful to the following third-party replicators (individuals[19] and institutions, in no particular order) who assisted us with verifications when we were unable to access data or, in some cases, computing resources: Peri Weisberg (San Francisco Human Services Agency), Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Majken Stenberg, CHRR at The Ohio State University, Florio Arguillas at the Results Reproduction (R-Squared) service at Cornell University, Kan Yao (Yale University), Yulia Zhestkova (University of Chicago), Kamen Velichkov (Wharton), Ian Schmutte (University of Georgia), Paulo Guimarães (Banco de Portugal Microdata Research Laboratory - BPLIM), Arpita Patnaik (University of Wisconsin), Joint Committee on Taxation of the Congress of the United States, Upjohn Institute (2x), Kevin McKinney (U.S. Census Bureau), Vítor Costa (Cornell University, 2x), Rahul Kasar and Andrea Garcia (Federal Reserve Board). We in particular want to thank Olivier Akmansoy, Christophe Hurlin (Université d'Orléans), and Christophe Pérignon (HEC Paris), all of cascad, a certification agency for scientific code and data, who have been generous of their time and resources, and have provided us with 21 reports. They have been able to leverage their access to confidential French administrative data (customs and labor market data), and have served as our proxies when data access was only available to European residents for Swedish administrative data.

We thank the following individuals who provided us with direct access to data and computing resources under their control, in order to prepare a reproducibility report. In such cases, we were able to conduct our checks ourselves, with full independence: Marc-Andreas Muendler (UCSD), Dan Feenberg (NBER).

We do not name the authors with whom we signed non-disclosure agreements, or who otherwise provided us with access to data that could not be published. We are grateful for their flexibility and patience.

---

[19]We do not name individuals when doing so would reveal information not already known to the manuscript's authors, naming the institution instead.