Report for 2021 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. We currently conduct pre-publication reproducibility checks for all AEA journals, and conduct basic checks on replication packages for Papers and Proceedings.

As in previous years, 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.

I. Providing Support for Compliance with AEA Data and Code Availability Policy

Since the introduction of the AEA's strengthened data and code availability policy in 2019 (American Economic Association, 2020), we have monitored how authors work to comply with the policy upon first submission of their packages. To help authors, we have published and continually update guidance available at aeadataeditor.github.io. The template README (Vilhuber et al., 2020), which we published with several other economics data editors, helps authors compile all the information required for complete documentation of their data and code deposit.[1] As more and more authors provide us with the standardized README, we have observed an uptick in replication packages that can be accepted (possibly with minor edits) in the first round (see Table 3). Compliance is not yet universal, despite the standard checklist provided and filled by all authors having a checkbox indicating self-asserted compliant use of the template README.

We have published guidance on specific topics in the form of blog posts[2], which are highlighted on the AEA Data Editor's Twitter account (@AeaData). Multiple presentations at workshops, conferences, and departmental seminars are also intended to clarify the AEA's policies on reproducibility, and convey best practices. Talks with materials are listed at aeadataeditor.github.io/talks/.

A. *Improving Findability of Replication Materials at the AEA*

We endeavor to make the replication packages provided by the authors broadly and easily findable. Naturally, they are linked from the article landing pages, but search engines such as Web of Science[3] and Google Scholar[4] can also lead interested researchers to the replication packages directly. Authors are invited to fill out rich metadata as appropriate for their paper upon submission.[5] For

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

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

[2] Blog posts by the AEA Data Editor can be found at aeadataeditor.github.io/year-archive/.

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

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

[5] See guidance on metadata at aeadataeditor.github.io/aea-de-guidance/data-deposit-aea.html.

deposits made prior to July 2019 and migrated to the AEA Data and Code Repository in October and December 2019, metadata was missing. In summer and fall of 2020, thanks to the support provided by a research team at the University of Michigan and ICPSR, we invited authors to update the metadata. 911 authors and provided additional metadata on 522 replication packages. After review by curation specialists at ICPSR, the expanded metadata was published for these deposits between 2021-08-26 and 2021-10-27, making it available to indexing services.

Authors are reminded that deposits receive their own Digital Object Identifier (DOI), and should be cited in line with the Data Citation Principles (Altman and Crosas, 2013; Data Citation Synthesis Group, 2014). We continue to verify that authors cite all datasets they have used and accessed, as required by the AEA data and code availability policy (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[6] 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.[7]

B. *Post-publication modifications*

At the time of publication, a manuscript is linked with one (or more) archived supplements, constituting the version of record. Occasionally, issues are brought to our attention by authors, readers, and data providers. Authors may have a better README, readers might have noticed a missing code or data file, or data providers might ask for a dataset to be removed that infringes on terms of use agreed to by the author. The supplemental "Policy on Revisions of Data and Code Deposits in the AEA Data and Code Repository[8]" specifies 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 identified 15 actions regarding post-publication modifications in 2021. Of these, 7 were reader-initiated, 4 were author-initiated, and 4 were initiated by the Data Editor. Of these, 3 were related to the posting of datasets that were not in compliance with the terms of use. All were brought into compliance with both the DCAP and any data provider terms of use, or are pending such resolution.

C. *Intellectual Property and Licenses*

Authors retain copyright for any data and code deposited by them in the AEA Data and Code Repository, unless that copyright belongs to others and the authors have a license to republish it. 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. All licenses are vetted by the Data Editor for compliance with the DCAP. We encourage authors to consult our licensing guidance.[9]

In a 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. Examples include Deryugina, Shurchkov and Stearns (2021*b*) and Goncalves and Mello (2021*b*),

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

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

[8] https://www.aeaweb.org/journals/data/policy-revisions

[9] https://aeadataeditor.github.io/aea-de-guidance/licensing-guidance

which accompany Deryugina, Shurchkov and Stearns (2021*a*) and Goncalves and Mello (2021*a*), respectively. In each case, the replication data contained sensitive data, which required an openly accessible but controlled method of distribution. Authors who wish to explore ways to make their data ethically accessible should contact the Data Editor early enough in the submission process.

The AEA replicators will sometimes access confidential or proprietary data for the purpose of verifying computational reproducibility (see Section II.A), as provided by the authors, or directly requested from the data providers via application or subscription services. Such data are not published as part of authors' replication packages. However, we do encourage authors to seek permission to share such data, where possible, and encourage data providers to allow for publication of extracts of their data, sufficient to support future reproducibility efforts.

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.

D. 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 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).

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. In general, the first two infrastructure pieces 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. In 2021, the Data Editor also piloted the use of several other infrastructures for conducting reproducibility checks and for the preservation of data for replication packages.

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 are 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 2). 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 publication.[10] However, to streamline processing, we may also recommend an "acceptance with modifications requested." In such cases, the remaining modifications are minor, and can be handled during copy-editing (for instance, a small number of tables need

[10]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.

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.

Assessments made

Between 2020-12-01 and 2021-11-30, the AEA Data Editor team received 529 requests, for 415 manuscripts.[11] 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, 490 reports (384 manuscripts) were submitted back to editors,[12] and 249 were completed up to the point of publication of the data deposit, including any post-acceptance modifications. Table 2 shows the distribution of the last recommendation on record for manuscripts as of 2021-11-30. Table 1 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.

[11] This includes only requests submitted between those dates, and does not take into account in-progress requests on 2020-12-01.

[12] 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.

Issues encountered

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.

Specification of computational environment: As we noted last year, we do not systematically encounter "manifest"-like files in R, Python, and Julia, even when such languages are used (which remains rare). Precise specification of third-party packages (in Stata, R, Python, and Julia) and of an accurate and useful description of the computational environment of the original researcher remains rare. No replication package explicitly made use of user-provided "containers" (docker, singularity), though two replication packages implicitly made use of these through CodeOcean, and we actively worked with some users to leverage such environments (see our discussion later under *Computational Infrastructure*).

Incomplete instructions and manual manipulation: Heuristically, the number of manual instructions to run code, or to save tables and figures, which detract from speedy and efficient reproduction by third parties, remain too high. However, we have noted an improvement in general in terms of data provenance description and of instructions for computations since the introduction of a template for a README (Vilhuber et al., 2020), which is shared by multiple journals.

Delays

A recurring concern expressed by authors, editors, and staff members are delays in publication, due to the verification process. The median manuscript is reviewed once (Table 3 shows the breakout by journal). In the previous year, the median manuscript went through two rounds of reviews. We have been able to reduce the number of

Table 1—: Processing Statistics

Journal	Issues (rcvd)	(cplt)	(external)	Manuscripts (rcvd)	(cplt)	(ext.)	(pend.)
AEJ:Applied Economics	119	111	1	78	73	1	39
AEJ:Economic Policy	119	108	4	92	85	4	48
AEJ:Macro	63	56	2	50	43	2	27
AEJ:Micro	39	38	1	29	29	1	20
AER	116	109	3	104	97	3	70
AER:Insights	30	28	NA	27	25	NA	17
JEL	9	7	NA	8	6	NA	4
JEP	34	33	NA	27	26	NA	24

Notes: Data for requests received by the AEA Data Editor between 2020-12-01 and 2021-11-30. AEA P&P are excluded from this table. See text for details.

Table 2—: Recommendations

Response option	Frequency
Accept	69
Accept - with Changes	290
Conditional Accept	12
Revise and Resubmit	13

rounds before a paper is accepted by accepting replication packages subject to minor post-acceptance edits (the "accept with changes" decision described earlier). Figure 1 illustrates the difference graphically, by journal.

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.[13] Each of these forms has links to updated and detailed guidance on how to prepare 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 first rolled out in September 2020, and revised in June 2021. In addition, a template "README" (Vilhuber et al., 2020) was introduced in December 2020 as a joint effort with other data editors, helping to clarify the necessary information, and to provide structure to authors' efforts to document their processes.

Computational Infrastructure

Most replication packages are computationally verified by replicators on the computers available to the Data Editor at the Cornell University Economics Department and the ILR School. The majority are handled on the Windows Server systems of the Cornell Center for Social Sciences, while some are run on the Linux-based Bioinformatics cluster. Occasionally, personal macOS laptops are used. Systems can handle memory requirements up to 1024 GB or up to 100

[13] These forms can also be found at aeaweb.org/journals/data.

Table 3—: Assessment rounds for completed manuscripts

Rounds	AER	AER Insights	AEJ Applied	AEJ Macro	AEJ Micro	AEJ Policy
1	69	18	37	15	14	57
2	11	3	9	13	8	14
3	0	0	0	0	1	1

Notes: Data for papers first sent to the AEA Data Editor between 2020-12-01 and 2021-11-30. AEA P&P, JEP, and JEL are excluded from this table. See text for details.

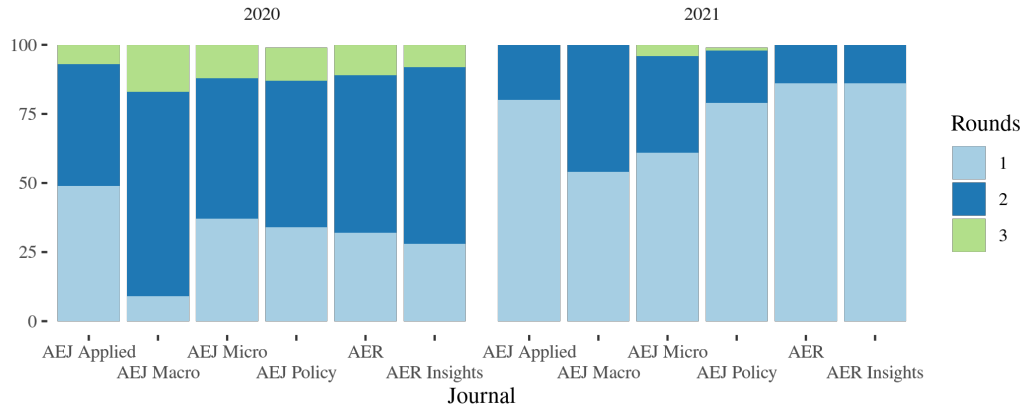

Figure 1. : Comparing rounds per journal between 2020 and 2021

cores.

While these systems are fairly standard, they cannot cover all scenarios described in authors' computational requirements. Furthermore, these systems, much like the authors' own systems, are not shareable more broadly, and thus sometimes make it difficult to control for specific requirements, or to share error messages in the most reproducible way.

Increasingly, we have therefore explored and documented additional computational environments. We have used CodeOcean,[14] (Clyburne-Sherin, Fei and Green, 2019) both to share active (but only partially successful) reproduction efforts and to publish reproducible "capsules." We have also explored the use of "WholeTale" (Chard et al., 2020), collaborating with its maintainers and other reproducibility institutions (e.g., the Odum Institute, which conducts reproducibility checks for various political science journals) to expand the utility of the system for the social sciences. Both CodeOcean and WholeTale rely on containerization, often known under the commercial name "Docker," which can be independently used to precisely define and then share computational environments. We have used containers directly in some instances, including with licensed software (Gurobi, Stata, geocod.io) and when data assets are too large to be accepted into either CodeOcean or WholeTale, running such containers on the Linux systems at Cornell University. Table 4 lists some recent examples. Sample code that is used by AEA replicators and can be used freely by any researcher can be found at the AEA Data Editor's Github repository.[15] A more expansive overview of containerization issues in economics can be found at AEA Data Editor (2021*c*).

We plan to continue exploring novel computational tools, their utility in accelerating and streamlining the verification process, and their availability for use for more efficient replication packages in economics.

B. Archive for Replication Packages

The default archive for replication packages accompanying articles in AEA journals is the AEA Data and Code Repository. Deposit instructions are provided on the Data Editor's website, and mentioned upon conditional acceptance. However, it is not the only acceptable archive, as we discuss below.

The default allowed package size is 30GB. In the past year, the median package size was 35.41 MB, but a significant number of packages (19 percent) had packages larger than 2GB. The move to a larger default size significantly reduced hurdles for authors. 3 percent of deposits were larger than 20GB. Some packages have more than 1,000 files, hitting a technical constraint. Provision of opaque ZIP files are generally prohibited. Instructions on how to proceed when file numbers are large, while maintaining maximum visibility onto the file and package structure, are provided on the website. Authors with large packages, or packages with more than 1,000 files, should contact the AEA Data Editor. Depositing at other trusted repositories is one option, described in the next section.

C. 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. 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. Only a few authors have availed themselves of this option, depositing or referencing materials in particular at the Harvard Dataverse[16] and Zenodo.[17] To support authors depositing on Zenodo, we have created a "Zenodo community" at zenodo.org/communities/aeajournals/. At this time, examples include Fohlin and Lu

[14] https://codeocean.com
[15] https://github.com/AEADataEditor/
[16] https://dataverse.harvard.edu/
[17] https://zenodo.org/

Table 4—: Use of containerization for replication packages

Replication package	Technology
Rossi (2021*a*,*b*)	CodeOcean, Docker, Matlab
DellaVigna and Pope (2021*a*,*b*)	CodeOcean, Docker, Stata
AEA Data Editor (2021*b*); Dinkelman and Ngai (2022)	Docker, Stata
AEA Data Editor (2021*a*); Brunnermeier et al. (2021)	Docker, R
Assunção, Gandour and Rocha (2021)	CodeOcean, R

Notes:

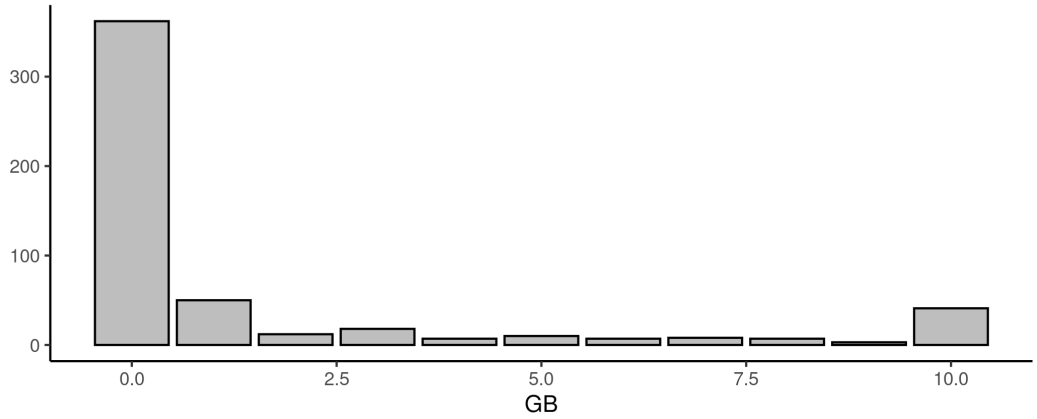

Figure 2. : Size distribution of replication packages deposited between 2020-12-01 to 2021-11-30, top-coded at 10GB.

(2021), who deposited code and data for their P&P paper, and Gendron-Carrier et al. (2020), whom we helped deposit data that was too cumbersome to deposit on ICPSR infrastructure (> 200GB).[18] Additional Zenodo deposits are currently being prepared for publication in 2022. Third-party repositories are linked to the main AEA Data and Code Repository deposit, and are cited in the main article when appropriate. Authors wishing to deposit replication packages early in the research lifecycle are encouraged to consult the Social Science Data Editors website,[19] where links to trusted repositories are provided.

[18]Code to support uploading large quantities of data to Zenodo via the Zenodo API, originally created by LDI Lab Member Vansh Gupta, can be found at github.com/AEADataEditor/Upload-to-Zenodo.

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

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 includes other publishers 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 and Preserving Data Resources

We identified Zenodo as a possible solution to preserving large-scale data resources. By creating the "AEA Zenodo Community," we are able to highlight data resources that have, in the past, often not been curated

or preserved appropriately. In addition to very large data resources, we also occasionally work to preserve data resources created by the U.S. government, used in multiple articles, but not formally preserved elsewhere. This includes NIH National Cancer Institute (2021); National Center for Chronic Disease Prevention and Health Promotion (2021*a*,*b*,*c*). As needed, these will be updated.

B. *Data Providers*

We regularly meet and communicate with academic, governmental, 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.

In the course of the past year, we have talked about some or all of these topics with IPUMS, the World Bank, Standard and Poor's, the Federal Reserve of St. Louis data librarians, John Abowd, Rob Sienkiewicz, and Barbara Downs (U.S. Census Bureau), Philipp vom Berge and Dana Müller at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Paulo Guimarães at the Banco de Portugal Microdata Research Laboratory (BPLIM), Ricardo Dahis (Base dos Dados, basedosdados.org), and Kamel Gadouche and Roxanne Silberman at the French *Centre d'accès sécurisé de données* (CASD). We have also talked to various research groups on how to improve data curation, visibility, and citability of the data created by their efforts.

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 to interact with others.[20] Mailing list members who wish to be more actively involved can participate in the development of the website of the Social Science Data Editors.[21] 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 authors of the template README, Lars Vilhuber, Miklos Kóren (Review of Economic Studies), Joan Llull (Economic Journal), Peter Morrow and Marie Connolly (Canadian Journal of Economics), continue to collect input on improvements on the README, and expect to have an updated and improved version available in 2022. The Data Editor has also been invited to the Steering Committee of a group of data repository leaders at Data-PASS organized as the Journal Editors Discussion Interface (JEDI[22]).

D. *Third-party verification services*

We continue to rely on and have discussions with third-party verification services. As noted earlier, 11 reports were provided by external replicators or replication services for 11 manuscripts (see Table 1 for statistics by journal, Appendix VI.A for a list of third-party replicators).

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

Outreach through presentations and publicly available tools is a key component of an effective data and code availability policy.

A. *Presentations*

The Data Editor participated in the AEA Committee on the Status of Women in the Economics Profession (CSWEP) "Fireside chats[23]" on January 19, 2021 with the topic "Demystifying Replication Requirements and Processes: Best Practices Viewed

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

[21] socialsciencedataeditors.github.io

[22] https://dpjedi.org/

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

by AEA Data Editor." Presentations on reproducibility in economics, and the various efforts surrounding the Data and Availability Policy, have been presented at the NSF Webinar on Best Practices in Making Research Datasets Publicly Accessible, the Research Data Alliance's 17th Plenary, a CNSTAT Expert meeting on Guidance on Data Sharing for NIA Longitudinal Studies, the Future of Privacy Forum Promoting Responsible Research Data Access, and at workshops at the IAB, CES-ifo, NBER, and Banco de Portugal. Recordings of presentations (when available) and presentations materials are listed at the Data Editor's website.[24] Together with BITSS (UC Berkeley), a framework and platform to support teaching reproducibility is being developed, called the "Social Science Reproduction Platform" (www.socialsciencereproduction.org), complementing other existing services such as the ReplicationWiki. The platform has been presented at CTREE 2021 and the ASSA 2022 meetings. A paper on the education and training of undergradudate replicators is under review at the *Journal of Statistics and Data Science Education* as of January 1, 2022.

B. *Resources*

The AEA Data Editor maintains public resources available to the economics community. These are made available through a dedicated website at aeadataeditor.github.io/ and code and project templates provided at github.com/aeadataeditor. In particular:

- Step-by-step guidance on how to prepare a replication package is provided at aeadataeditor.github.io/aea-de-guidance/, including video tutorials and a description of the process.
- The template README (Vilhuber et al., 2020) is referenced as part of the guidance, and separately accessible at social-science-data-editors.github.io/template_README/.
- Various blog posts on topics relating to computational reproducibility are posted at aeadataeditor.github.io/year-archive/ and typically summarized on Twitter under the Data Editor's handle @AEAData.
- Instructions to replicators for assessing authors' replication packages are provided at github.com/AEADataEditor/replication-template.
- Template code for using containers for Stata, R, Julia, and Gurobi can be found by searching for "docker" on the Github site.

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

V. Replication team at Cornell University

A. *Replicators*

The following **44** students have provided excellent assistance in reproducing the results from the 415 articles processed by the Replication Lab: Andreas Psahos, Asha Patt, Ashley Cooray, Craig Schulman, Daniella Pena, Dmitry Shlyapnikov, Eli Kolodezh, Emma Sbrollini, Hongyi Duan, Huey Le, Jacob Recht, Jaeyoung Shim, Janet Malzahn, Jared Martin, Jill Crosby, Jonathan Temkin, Joshua Passell, Julia Zimmerman, Kate Hofer, Kevin Bao, Liam Cushen, Lilly Thomalla, Lincy Chen, Lydia Reiner, Matt Wang, Matthew LaFontaine, Melanie Brown, Melanie Chen, Miranda Zhou, Nehedin Juarez, Ololade Omotoba, Qianyi Liu, Ryan Ali, Sam Evans, Satya Datla, Steve Yeh, Surita Basu, Suvd Khaliun, Tarangana Thapa, Taren Daniels, Vansh Gupta, Xiangru Li, Zebang Xu, Zechariah Karsana.

Graduate students Hyuk Harry Son and Leonel Borja Plaza and Research Aide Michael Darisse (all 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. Leonel Borja Plaza contributed programming to this report.

B. *Computing support*

We thank the Economics Department and the ILR School for providing us with computing resources at the Cornell Center for

Social Sciences and the Bioinformatics cluster.

VI. Third-party contributors

A. Replicators

We are grateful to the following third-party replicators, who assisted us with verifications when we were unable to access data or, in some cases, computing resources. We do not name individuals when doing so would reveal information not already known to the manuscript's authors, naming the institution instead. Names are listed in no particular order. Daniel Feenberg (National Bureau of Economic Research (NBER)), Karma Sonam and Keenan Dworak-Fisher (both Bureau of Labor Statistics (BLS)), Paulo Guimarães (BPLIM), Lisa Neilson and Joshua Hawley (Ohio Longitudinal Data Archive (OLDA)), Philipp vom Berge (IAB), Ian Schmutte (UGA) graduate students Caitlin Ahearn (UCLA), Manuel V. Montesinos (Universitat Autònoma de Barcelona, on behalf of the Data Editor of the Economic Journal, Joan Llull) We in particular want to again 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 4 reports during this time.

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.

B. Computing resources

We are grateful to CodeOcean, NBER, WholeTale, and Harvard Business School, who all provided us with access to computing resources at no cost, and technical assistance when necessary. We use free academic resources on Github and Bitbucket.

VII. Disclosures

We received a generous compute and storage quota from CodeOcean, a free license to use Stata 17 for one year in cloud applications from Stata, and a subaward on NSF grant 1541450 "CC*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale" from the University of Illinois to evaluate the WholeTale platform for the purpose of reproducibility verification. None of the sponsors have reviewed this preliminary assessment, or have had influence on any of the conclusions of this document. CodeOcean currently offers academic users a certain number of monthly free compute hours. WholeTale is free to use, the Stata functionality mentioned here is expected to be available in early 2022.

VIII. Data and Code Availability Statement

All publicly available data and code used to generate figures and tables in this article are available (Vilhuber, 2022*a*,*b*). Some detailed data from the editorial system, used for Table 3, are considered confidential and cannot be made available in a way that preserves the privacy of the editorial process at this time.

LARS VILHUBER, *Data Editor*

REFERENCES

AEA Data Editor. 2021*a*. "AEADataEditor/Docker-Aer-2018-0733: Docker Image." *AEA Data Editor Github repository*. https://github.com/AEADataEditor/docker-aer-2018-0733 (accessed 2022-01-03).

AEA Data Editor. 2021*b*. "AEADataEditor/JEP-2021-1239: Docker Image." *AEA Data Editor Github repository*. https://github.com/AEADataEditor/JEP-2021-1239 (accessed 2022-01-03).

AEA Data Editor. 2021*c*. "Use of Docker for Reproducibility in Economics." https://aeadataeditor.github.io/posts/2021-11-16-docker (accessed 2022-01-03).

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.

Assunção, Juliano, Clarissa Gandour, and Romero Rocha. 2021. "Replication Capsule for: DETERring Deforestation in the Amazon: Environmental Monitoring and Law Enforcement [Source Code]." CodeOcean, https://doi.org/10.24433/CO.5098352.v1.

Brunnermeier, Markus, Darius Palia, Karthik A. Sastry, and Christopher A. Sims. 2021. "Feedbacks: Financial Markets and Economic Activity." *American Economic Review*, 111(6): 1845–1879. https://doi.org/10.1257/aer.20180733.

Chard, Kyle, Niall Gaffney, Mihael Hategan, Kacper Kowalik, Bertram Ludaescher, Timothy McPhillips, Jarek Nabrzyski, Victoria Stodden, Ian Taylor, Thomas Thelen, Matthew J. Turk, and Craig Willis. 2020. "Toward Enabling Reproducibility for Data-Intensive Research Using the Whole Tale Platform." *arXiv:2005.06087 [cs]*. https://doi.org/10.3233/APC200107.

Clyburne-Sherin, April, Xu Fei, and Seth Ariel Green. 2019. "Computational Reproducibility via Containers in Psychology." *Meta-Psychology*, 3. https://doi.org/10.15626/MP.2018.892.

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.

DellaVigna, Stefano, and Devin Pope. 2021*a*. "Compute Capsule for: Stability of Experimental Results: Forecasts and Evidence [Source Code]." CodeOcean, https://doi.org/10.24433/CO.0687784.v1.

DellaVigna, Stefano, and Devin Pope. 2021*b*. "Data and Code for: Stability of Experimental Results: Forecasts and Evidence." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E135221V1.

Deryugina, Tatyana, Olga Shurchkov, and Jenna Stearns. 2021*a*. "COVID-19 Disruptions Disproportionately Affect Female Academics." *AEA Papers and Proceedings*, 111: 164–168. https://doi.org/10.1257/pandp.20211017.

Deryugina, Tatyana, Olga Shurchkov, and Jenna Stearns. 2021*b*. "Data for: COVID-19 Disruptions Disproportionately Affect Female Academics." American Economic Association [publisher] Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E139263V1.

Dinkelman, Taryn, and L. Rachel Ngai. 2022. "Data and Code for: Time Use and Gender in Africa in Times of Structural Transformation." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E145161V2.

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.

Fohlin, Caroline, and Zhikun Lu. 2021. "Code for "How Contagious was the Panic of 1907? New Evidence from Trust Company Stocks"." Zenodo, https://doi.org/10.5281/zenodo.5151203.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A. Turner. 2020. "Global daily Aerosol Optical Depth measurements from Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA's Aqua and Terra satellites." Zenodo, https://doi.org/10.5281/zenodo.4317553.

Goncalves, Felipe, and Steven Mello. 2021*a*. "A Few Bad Apples? Racial Bias in Policing." *American Economic Review*, 111(5): 1406–1441. https://doi.org/10.1257/aer.20181607.

Goncalves, Felipe, and Steven Mello. 2021*b*. "Supplementary Data for: A Few Bad Apples? Racial Bias in Policing." American Economic Association [publisher] Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E123921V1.

National Center for Chronic Disease Prevention and Health Promotion. 2021*a*. "Behavioral Risk Factor Surveillance System (BRFSS) Annual Survey Data, 1999-2015." Centers for Disease Control and Prevention [publisher] Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E146342V1.

National Center for Chronic Disease Prevention and Health Promotion. 2021*b*. "Behavioral Risk Factor Surveillance System (BRFSS) Annual Survey Data, 1999-2019." Centers for Disease Control and Prevention [publisher] Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E146342V2.

National Center for Chronic Disease Prevention and Health Promotion. 2021*c*. "SMART: BRFSS City and County Data (2002-2015)." Centers for Disease Control and Prevention [publisher] Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E146343V1.

NIH National Cancer Institute. 2021. "NIH SEER U.S. Population Data - 1969-2019." National Institutes of Health [publisher] Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E146341V1.

Rossi, Barbara. 2021*a*. "Compute Capsule for: Forecasting in the presence of instabilities: How do we know whether models predict [Source Code]." CodeOcean, https://doi.org/10.24433/CO.7940775.v1.

Rossi, Barbara. 2021*b*. "Data and Code for: Forecasting in the Presence of Instabilities." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E147225V1.

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. 2022*a*. "Code and Data for: Report for 2021 by the AEA Data Editor." American Economic Association [publisher], https://doi.org/TBD.

Vilhuber, Lars. 2022*b*. "Process data for the AEA Pre-publication Verification Service." American Economic Association [publisher] V3, https://doi.org/10.3886/E117876V3.

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.