Reproducibility and Replicability in Economics

Lars Vilhuber[1,2]

[1]Labor Dynamics Institute, School of Industrial and Labor Relations, Cornell University, Ithaca, New York, United States of America,
[2]Department of Economics, The College of Arts and Sciences, School of Industrial and Labor Relations, Cornell University, Ithaca, New York, United States of America

ABSTRACT

I provide a summary description of the history and state of reproducibility and replicability in the academic field of economics. I include a discussion of more general replicability and transparency, including the tradition of sharing of research findings and code outside of peer-reviewed publications. I describe the historical context for journals and grey literature in economics, the role of precollected public and nonpublic data, and touch on the role of proprietary software in economics. The increasing importance of restricted-access data environments in economics and the interaction with reproducibility is highlighted. The article concludes with an outlook on current developments, including the role of big data and increased verification of reproducibility in economics.

Keywords: reproducibility, transparency, replicability, preprints, open-source, journal policies

1. Introduction

In this overview, I provide a summary description of the history and state of reproducibility and replicability in the academic field of economics. I will attempt to discuss not just the narrower definition of computational reproducibility but also other correlates of intellectual reproducibility and transparency, such as the sharing of research findings outside of peer-reviewed publications ('grey publications'), and the importance of various types of data for empirical economics.

I start by defining reproducibility and replicability. Our focus is primarily on the journals that are the prime publication outlets of academic economists, and the role they have and can continue to play. Part of the reason for this focus is because it is much easier to measure replicability for published materials (even if what is being measured may change from study to study). Nevertheless, the informal and non-peer-reviewed sharing of documents, code, and data plays an important role in economics. I describe the historical context for journals and grey literature, review the historical roots of the use of precollected public and nonpublic data, and touch on the role of proprietary software in economics. This discussion frames the description of the state of reproducibility and replicability in modern economics, which here means within the last 30 years. I highlight the increasing importance of restricted-access data environments in economics and the interaction with reproducibility. In contrast, the role of replication, reproduction, and emulation in the teaching of economics is much harder to assess, though I will provide some indications as to its use in education. I then describe what is currently occurring in economics, touching on topics like big data and reproducibility in economics, or the search for the right method to surface reproductions and replications. Much of this is new, and the evidence on sustainability and impact is yet to be collected.

The purpose of the overview is not to propose specific solutions, but rather to provide the context for the multiplicity of innovations and approaches that are currently being implemented and developed, both in

economics and elsewhere. This overview will focus almost exclusively on economics, but should be read in conjunction with other articles in this issue on the same topic, which in turn describe the same topic in other disciplines. All were originally written to inform the National Academies' Committee on Reproducibility and Replication in Science in preparing their 2019 report (National Academies of Sciences, Engineering, and Medicine [NASEM], 2019). Finally, I have found the knowledge gained from compiling this overview useful in clarifying my vision for the data and code policy at the American Economic Association (AEA), and map out a feasible but ambitious roadmap for improving reproducibility at the AEA's multiple journals (Vilhuber, 2019; Vilhuber et al., 2020). It may also be useful for others in similar positions, who may endeavor to ameliorate a discipline's reproducibility and transparency practices.

2. Definitions

In this text, we adopt the definitions of *reproducibility* and *replicability* articulated, inter alia, by Bollen et al. (2015) and in the report by NASEM (2019). In economics, as in other sciences, a variety of usages and gradations of the terms are in use (Clemens, 2017; Hamermesh, 2007, 2017; *Journal of Applied Econometrics*, 2014). At the most basic level, *reproducibility* refers to "to the ability […] to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator" (Bollen et al., 2015, p. 3). "Use of the same procedures" may imply using the same computer code, or reimplementing the statistical procedures in a different software package, as made explicit in the notion of *narrow replicability* (*Journal of Applied Econometrics*, 2014; Pesaran, 2003). Reproducibility may be seen as analogous to a 'unit test'[1] in software engineering.

Replicability, on the other hand, refers to "the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected" (Bollen et al, 2015, p. 4), and *generalizability* refers to the extension of the scientific findings to other populations, contexts, and time frames. Because there is a grey zone between these two definitions, we will generally refer to either context as 'replicability.' Hamermesh (2007) calls this "scientific replication." Robustness tests performed by researchers have aspects of self-replication, by identifying conditions under which the findings continue to hold when software or data are varied.

In this article, we will use the terms as defined above, even when authors use different terms.[2]

3. Historical Context

3.1. Replicability and Reproducibility in Early Economics

Publication of research articles specifically in economics can be traced back at least to the 1844 publication of the *Zeitschrift für die Gesamte Staatswissenschaft* (Stigler et al., 1995). The AEA was founded in 1885, though initially most articles were not novel research reports (Margo, 2011). U.S.-based journals that were founded at the time were Harvard's *Quarterly Journal of Economics* (1886) and the University of Chicago's *Journal of*

Political Economy (1892), and in the United Kingdom, the Royal Economic Society founded the *Economic Journal* in 1891 (Stigler et al., 1995). However, publications by prominent economists had appeared in generalist academic journals prior to those initial issues (Stigler et al., 1995). The modern-day *American Economic Review* (*AER*) followed in 1911 and the *Review of Economics and Statistics* in 1918. Of some significance in the context of replicability was the founding of *Econometrica* in 1933. As the first editor of *Econometrica*, Ragnar Frisch noted, "the original data will, as a rule, be published, unless their volume is excessive […] to stimulate criticism, control, and further studies" (Frisch, 1933). Most data at the time would have been published in paper form, and cited as such, as there would have been no distinction between 'data' and 'text' as we generally observe it today. However, editors in later years of *Econometrica*, as well as of the other journals, put rather less emphasis on this aspect of the publication process. Whether this is due to specialization—only 17.4% of articles in *Econometrica* in 1989–1990 had empirical content (Stigler et al., 1995)—or for other reasons, is unknowable.

Much of economics was premised on the use of statistics generated by national statistical agencies as they emerged in the late 19th and early 20th century.[3] These were already so prevalent as a source of data, to be used in economic research and broadly sharable and shared, that the founding issue of the *Review of Economics and Statistics* explicitly precluded duplicating such collection and dissemination of data (Bullock, 1919). At the same time, data sharing was easier: the same founding issue simply published tables of data as used by the author, both "original" and "computed" ("Standard Charts and Tables," 1919). There was, it can be argued, a greater similarity of 'reproducibility' between theoretical economics (where proofs can be verified) and applied economics (where manual calculations, given the printed data, can be verified).

The emphasis on statistical and empirical analyses increased, not just in the *Review of Economics and Statistics* but also in the more technical *Econometrica* and the more generalist *American Economic Review* (Margo, 2011). By the late 1950s, the idea of even greater access to published and confidential government data by a large group of 'data users,' including economists, was well accepted (ASA Advisory Committee, 1959; Kraus, 2013). This same period also saw the creation of archives, such as the Inter-University Consortium for Political Research (soon renamed as the Inter-University Consortium of Political and Social Research, ICPSR), specifically designed to collect, convert, standardize, and disseminate electronic records to academics, from surveys and other sources (ICPSR, n.d.; Miller, 2016).[4] Constraining wider dissemination was the ability to actually perform machine-based computations, as the Census Bureau and a few big universities were the only ones with sufficient compute power to actually leverage many of these data.

A key takeaway is that much of economic research relied on publicly available data. Initially, the data was just another form of (paper) publication, and thus easily identified and referenced by standard bibliographic citations. Later, with the advent of electronic records, a relatively small set of consortia and data providers were responsible for dissemination, by tapes, CD-ROMs and, starting in the 1990s, FTP servers. While this should lead to relatively unambiguous data citations, this seems to not have been the case. As Dewald et al. (1986)

note: "Many authors cited only general sources such as Survey of Current Business, Federal Reserve Bulletin, or International Financial Statistics, but did not identify the specific issues, tables, and pages from which the data had been extracted."

3.2. A History of Sharing Pre-Prints and Code

Economics has a history of sharing 'grey literature'—documents such as technical reports, working papers, and so on, that are typically not subject to peer-review (Rousseau et al., 2018), but are of sufficient quality that they are worthwhile preserving (Schöpfel, 2011), and in particular, worthwhile citing. Most scientists, when they think of preprints, think of arXiv (Ginsparg, 1997; Halpern, 1998), founded in 1991. However, the National Bureau of Economic Research (NBER) working paper series, one of the most prestigious in economics, was first published (in paper form) in 1973 (F. Welch, 1973). By the early 1990s, there was a wide variety of such working paper series, typically provided by academic departments and research institutions. Since grey literature at the time was not cataloged or indexed by most bibliographic indexes, a distinct effort to identify both working papers and the novel electronic versions grew from modest beginnings in 1992 at Université de Montréal[5] and elsewhere into what is today known as the Research Papers in Economics (RePEc) network (Bátiz-Lazo & Krichel, 2012; Krichel & Zimmermann, 2009), a "collaborative effort by hundreds of volunteers in 99 countries" (*RePEc: Research Papers in Economics*, n.d.). The initial index was split into electronic (WoPEc) (Krichel, 1997) and printed working papers (BibEc) (Cruz & Krichel, 2000; Krichel & Zimmermann, 2009), testimony to the prevalence of the exchange of scientific research in semi-organized ways. Economists had, in fact, access to a central repository for submitting working papers, based on the arXiv system, but it seems to not have been very popular, in contrast to the decentralized working paper archives (Krichel, 1997). In 1997, BibEc counted 34,000 working papers from 368 working paper series (BibEc, 1997). RePEc today has data from around 4,600 working paper series and claims about 2.5 million full-text (free) research items, provided in a decentralized fashion by about 2,000 archives (*IDEAS/RePEc*, n.d.). These items not only include traditional research papers, but also, since 1994, computer code (CodEc, 1998; *Economics Software | IDEAS/RePEc*, n.d.; Eddelbüttel, 1997). Although still cataloging mostly grey literature, RePEc bibliographic metadata is, in fact, indexed by all major bibliographic indexes.

3.3. The Increasing Importance of Nonpublic Data

Economists have been using nonpublic data that they have not themselves collected at least as far back as Adam Smith's pin factory.[6] Economists were requesting access for research purposes to government microdata through various committees at least as far back as 1959 (Kraus, 2013). Whether using private-sector data, school-district data, or government administrative records, from the United States and other countries, the use of these data for innovative research has been increasing in recent years. In 1960, 76% of empirical AER articles used public-use data.[7] By 2010, 60% used administrative data, presumably none of which is public use (see Figure 1). We will return to the effects of this phenomenon on reproducibility later, when discussing the effect of 'data policies.'
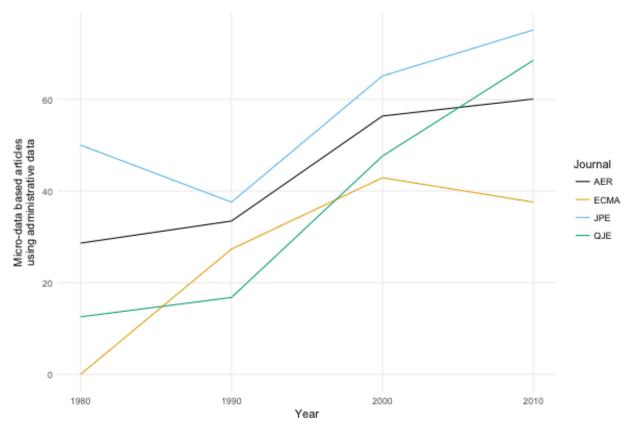
Figure 1. Use of administrative data in publications in leading journals, 1980–2010.
'Administrative' data sets refer to any data set that was collected without directly surveying individuals (e.g., scanner data, stock prices, school district records, social security records). Sample excludes studies whose primary data source is from developing countries. Figure reproduced from Chetty (2012) with permission.

3.4. Proprietary Software

Software is considered an important component of the reproducibility 'package.' Many economists have long been willing to (informally) share their custom code,[8] even if others are hesitant to do so. However, underlying this is a large dispersion in software tools, extending from Fortran 77 code to software instructions for popular (and typically proprietary) statistical software such as Minitab, SAS, SPSS, and Stata (released in 1985 for PCs). In particular, Stata is very popular among economists. Among reproducibility supplements posted alongside articles in the AEA's journals between 2010 and 2019, Stata is the most popular (72.96% of all supplements), followed by Matlab (22.45%; Vilhuber et al., 2020). Stata very soon had many of the trappings of today's open-access toolkit. The *Stata Journal*, where peer-reviewed add-ons for Stata are published, has a paywall, but the underlying programs can be installed for free and in source-code form by any user of Stata.[9] Additional open software archives are widely used and referenced (postings to Statalist since 1994, Statistical Software Components [SSC] archive[10] since 1997). Historically, software such as R (R Core Team, 2000), Python, and Julia (Bezanson et al., 2017) have not been widely used by economists, although each has had an

active economics community (Sargent & Stachurski, 2017). Figure 2 shows usage of various software programs within supplements at the AEA through 2019 (Vilhuber et al., 2020; Vilhuber, 2020).
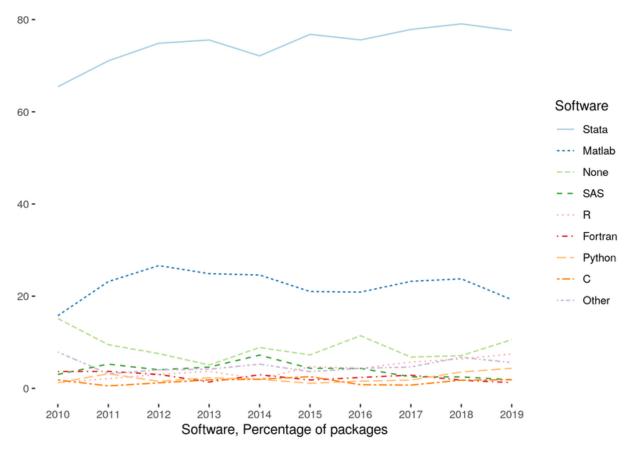

Figure 2. Software usage in supplements to journals of the American Economic Association, based on parsing of file extensions. From Vilhuber (2020)

4. Reproducibility and Replicability in Modern Economics

It is generally argued that the ability to replicate and validate scientific findings is an important, even critical part of the scientific method. When reproducing results, researchers can check for inadvertent errors, and code and data archives provide a basis for subsequent replications and extensions by others (King, 1995). In economics, complaints about the inability to properly conduct reproducibility studies, or about the absence of any attempt to do so by editors, referees, and authors, can be traced back to comments and replies in the 1970s (see (Dewald et al., 1986, for examples). Calls for better journal policies to support replicability were made (Feige, 1975). While the *Journal of Political Economy (JPE)* added a section to the journal for "verifications and contradictions" of papers published in the *JPE* between 1976 and 1987, this seems to not have been effective: only 36 notes were published, of which five were actually reproductions (McCullough et al., 2006), while the others were "employing either new data sets or alternative statistical techniques" (Dewald et al., 1986, p. 589). The best-cited example was the imposition of a "data availability policy" by the *Journal of*

Money, Credit, and Banking (JMCB). The subsequent analysis thereof (Dewald et al., 1986) considered all papers published, accepted, or under review by the *JMCB* between 1980 and 1984, of which some had been published before the announcement of the new data policy in 1982. The results suggested several problem areas. Authors, even among those whose article was still under review, had lost the data, or did not respond to the request for data and code. The nonsubmission rate was 65% for articles published before the announcement of the data policy, 26% after the announcement. Resource constraints (and the complexity of undertaking some of the replications) led to only 8 replication attempts being made, of which 5 were successful.[11] Only a few such systematic replication or reproducibility attempts were made in subsequent years. It was concluded that "there is no tradition of replication in economics" (McCullough et al., 2006).

4.1. Journal Policies Supporting Reproducibility

In the early 2000s, as in other sciences (National Research Council, 2003), journals started to implement "data' or 'data availability' policies. Typically, they required that data and code be submitted to the journal, for publication as 'supplementary materials.' The *JMCB* had reimplemented their policy in 1996, after a brief hiatus, and the *Journal of Applied Econometrics* has a data archive going back to 1988. The AEA announced its 'data availability policy' in 2003, implemented it in 2004, and extended it to the new domain-specific journals in 2009–2012. The first data supplements appear in *Econometrica* in 2004. The *JPE* announced its policy in 2004 and implemented it in 2005 (see Table 1 for details and links). Depending on how the sample of journals is selected, between 8.1% and 29.5% of economics journals (Duvendack et al., 2015; Vlaeminck & Herrmann, 2015) have a 'data availability policy.'[12]

Table 1. **Journal policies**.

Journals	Type of Policy	URL	Archive	Original Date	Notes
AER and Journals	AEA — Data and programs	https://www.aeaweb.org/journals/policies/data-availability-policy	Journal website	2004	Extended in 2009, 2012 to all journals of the AEA
Quarterly Journal of Economics	AEA	https://academic.oup.com/qje/pages/Data_Policy	Dataverse	2016	Non-compliance can lead to editorial expression of concern

Review of Economic Studies	Own — Data and Programs	https://academic.oup.com/restud/pages/General_Instructions	Journal website		
Review of Economics and Statistics	Own — Data and Programs	https://www.mitpressjournals.org/journals/rest/sub	Dataverse		Non-compliance leads to ban from submission for 5 years
Journal of Applied Econometrics	Own — Data only	https://onlinelibrary.wiley.com/page/journal/10991255/homepage/forauthors.html	Own (Queens, 1988–)	1988	Only data is compulsory.
Econometrica	Own — Data and Programs	https://www.econometricsociety.org/publications/econometrica/information-authors/editorial-procedures-and-policies#replication	Journal website	2004?	
Journal of Political Economy	AEA	https://www.journals.uchicago.edu/journals/jpe/datapolicy	Journal website	2005	
Journal of Money, Credit, and Banking	Own (barebones) — data only	https://jmcb.osu.edu/submission-instructions	Journal website	1982	Hiatus from 1993–1996.
Economic Journal	Own — Data and Programs	http://www.res.org.uk/view/datapolicyEconomic.html	Journal website	2013?	

The policies of the top journals listed in Table 1 generally reflect the lessons learned from earlier experiences. Although typically called 'data availability' policies, they are more accurately described as 'data and code availability' policies. Most attach an archive of both data and programs as 'supplementary data' on the journal website, with only the American Economic Association's journals and two Harvard-based journals (*QJE,*

ReStat) depositing materials in "proper" data archives—the former since 2019 using a repository hosted by ICPSR, and the latter on the (Harvard-based) Dataverse (n.d.). A consequence of this treatment of data supplements as secondary digital objects is that they do not generally obtain their own digital identifiers, the exception being those data supplements stored on archives. Few allow the data to be independently explored or discovered, with users being limited to obtaining the opaque ZIP files via links on journal websites. Few, if any, articles during this time period cite the data, even when the data and code objects have citable identifiers.

Journals in economics that have introduced data deposit policies tend to be higher ranked even before introducing the more stringent policy (Höffler, 2017b), possibly biasing analyses that focus on high-ranked journals (Crosas et al., 2018). None of the journals in Table 1 request that the data be provided before or during the refereeing process,[13] nor does a review of the data or code enter the editorial decision, in contrast to other domains (Stodden et al., 2013). All make provision of data and code a condition of publication, unless an exemption for data provision is requested.

Other journals have taken a more low-key approach, only requesting that authors provide data and code upon request postpublication. Studies old and new have found that the probability of obtaining sufficient data and code to actually attempt a reproduction is lower when no formal data or code deposit policy is in place (Chang & Li, 2017; Dewald et al., 1986; Stodden et al., 2018). In a simple check we conducted in 2016, we emailed all 117 authors that had published in a lower ranked economics journal between 2011 and 2013 (Vilhuber, 2020). The journal has no data deposit policy, and only requires that authors promise to collaborate. We sent a single request for data and code. Only 48 (41%) responded, in line with other studies of the kind (Stodden et al., 2018), and of those, only 12 (10% of total requests) provided materials upon first request.[14] Others report response rates of 35.5% (Dewald et al., 1986) and 42% (Chang & Li, 2017), with different request protocols and article selection criteria in each case.

More recently, economics journals have increased the intensity of enforcement of their policies. Historically being mainly focused on basic compliance, associations that publish journals, such as the aforementioned AEA, the Canadian Economics Association, the Royal Society (UK), and the association supporting the *Review of Economic Studies*, have appointed staff dedicated to enforcing various aspects of their data and code availability policies (AEA, 2019; Duflo & Hoynes, 2018; Vilhuber, 2019) . The enforcement varies across journals, and may include editorial monitoring of the contents of the supplementary materials, reexecution of computer code (verification of computational reproducibility), and improved archiving of data.[15]

4.2. Reproducibility Studies

If the announcement and implementation of data deposit policies improve the availability of researchers' code and data (R. G. Anderson & Dewald, 1994; Dewald et al., 1986), what has the impact been on overall reproducibility? A journal's data deposit policy needs to be enforced and verified—absence thereof can still

lead to low data and code availability: the nonsubmission rate among the 193 *JCMB* articles studied by McCullough et al. (2006) was 64%, even after the policy was theoretically in place (see Table 2A).

Table 2B, shows the reproduction rates both conditional on data availability as well as unconditionally, for a number of reproducibility studies.[16] In our own analysis, as well as in McCullough et al. (2006), a census of all articles over a certain time period was undertaken, whereas Chang & Li (2017) and Camerer et al. (2016) selected specific articles under certain search criteria. The studies undertaken in the past 3 years find a higher conditional reproduction rate than McCullough et al. (2006) (between 49% and 61%).

Table 2A. Submission and reproduction rates. Submission rates.

	Number of articles (requests)	Non-submissions	Confidential data	Non-submission rate	Non-submission excluding confidential data
Dewald et al (1986) before policy change	62	40	2	64.5%	63.3%
Dewald et al (1986) after policy change	92	24	1	26.1%	25.3%
McCullough et al (2006)	193	124	?	64.2%	n.a.
Chang and Li (2017)	67	21	6	31.3%	24.6%
Own analysis (AEA:AJ 2009-2013)	157	63	63	40.1%	0.0%

Note: For McCullough et al, we count as a submission if any data or code was submitted.

Table 2B. Submission and reproduction rates. Reproduction rates.

	Number of articles (requests)	Attempted reproductions	Successful reproductions	Reproduction rate | attempted	Reproduction rate | empirical article

Dewald et al. (1986) before policy change	62	5	3	60.0%	4.8%
Dewald et al. (1986) after policy change	92	3	2	66.7%	2.2%
McCullough et al. (2006)	193	62	14	22.6%	7.3%
Chang & Li (2017)	67	59	29	49.2%	43.3%
Own analysis (American Economic Journal: Applied Economics 2009-2013)	157	94	46	48.9%	29.3%
Camerer et al. (2016)	n.a.	18	11	61.1%	n.a.

4.3. The Importance and Impact of Restricted-Access Data

As the increase of nonpublic data in Figure 1 suggests, however, even if compliance among users of public-use data increases, it is possible for overall availability, and thus reproducibility, to decline. In the journals of the *AEA,* all authors complied with the policy, as evidenced by the various "Reports of the Editor" published each year by the *AEA* (Duflo, 2018; Goldberg, 2017), an improvement on earlier years (Glandon, 2011; Moffitt, 2011). However, as noted earlier, exemptions are given when restricted-access data is used in an article. In our analysis of all 157 articles appearing in *American Economic Journal: Applied Economics (AEJ:AE)* between 2009 and 2013, only 60% of articles have some data available—a lower percentage than in the original *JCMB* study (Table 2A). Note that exemptions are not clearly published or posted, and because all such papers are still required to provide the code used to process the confidential or proprietary data, most such papers still have a supplementary material ZIP file, but without data.

Data that is not provided due to licensing, privacy, or commercial reasons (often incorrectly collectively referred to as 'proprietary' data[17]) can still be useful in attempts at reproduction, as long as others can reasonably expect to access the data. For instance, while confidential data provided by the Health and Retirement Survey (HRS) or through the U.S. Federal Statistical Research Data Center (FSRDC) cannot be posted to journal websites, hundreds, if not thousands, of researchers have gained secure access to the same

confidential source data over the years, and could potentially reproduce or replicate the published research.[18] We thus analyzed each of the papers in the *AEJ:AE* that did not provide data, and classified the data used into five categories. Administrative data could be provided by a 'national' provider (a national statistical office or similar), a 'regional' entity (a state or province), or a 'local' entity (a school district, county, or other governmental institution). Private providers might be commercial (data for which access can be purchased, such as from Dun and Bradstreet, State Street, or Bureau van Dijk), or some other type. Table 3A, tabulates the distribution of characteristics among the 2009–2013 *AEJ:AE* articles with nonpublic data (Kingi et al., 2018). National data providers dominate in this journal, providing nearly 50% of all nonpublic data.

Providers will differ in the presence of formal access policies, and this is quite important for reproducibility: only if researchers other than the original author can access the nonpublic data can an attempt at reproducibility even be made, if it at some cost. We made a best effort to classify the access to the confidential data, and the commitment by the author or third parties to provide the data if requested. For instance, a data curator with a well-defined, nonpreferential data access policy would be classified under 'formal commitment.' The FSRDC or the German Research Data Center (FDZ) of the Institute for Employment Research (IAB) have such policies. If the author personally promises to provide access to the data, we further distinguished 'with commitment,' where the author would engage a third party to provide access in a well-defined fashion, from 'no commitment,' where the author would simply promise to work with a replicator, without being able or willing to guarantee such access. Our ability to make this classification depends critically on information provided by the authors. Table 3B, tabulates the results from that exercise. We could identify a formal commitment or process to access the data only for 35% of all nonpublic data sets.

Table 3A. Characteristics of nonpublic data. Type of non-public data.

Year	N	Local adminis tration	Nationa l adminis tration	Regiona l adminis tration	Private commer cial	Private other	Local adminis tration	Nationa l adminis tration	Regiona l adminis tration	Private commer cial	Private other
2009	4	0	2	1	0	1	0.0%	50.0%	25.0%	0.0%	25.0%
2010	17	2	8	0	4	3	11.8%	47.1%	0.0%	23.5%	17.6%
2011	16	2	9	4	1	0	12.5%	56.3%	25.0%	6.3%	0.0%
2012	15	1	10	2	0	2	6.7%	66.7%	13.3%	0.0%	13.3%
2013	11	2	2	1	4	2	18.2%	18.2%	9.1%	36.4%	18.2%
Total	63	7	31	8	9	8	11.1%	49.2%	12.7%	14.3%	12.7%

Note. Based on a manual classificaton of the 63 articles identified as having nonpublic data in the AEJ:AE in the years listed.

Table 3B. Characteristics of nonpublic data. Access.

Year	Formal commitment	Informal commitment	Informal, no commitment	No info		Formal commitment	Informal commitment	Informal, no commitment	No info
2009	4	0	0	0	4	100.0%	0.0%	0.0%	0.0%
2010	2	3	9	3	17	11.8%	17.6%	52.9%	17.6%
2011	3	0	10	3	16	18.8%	0.0%	62.5%	18.8%
2012	12	1	0	2	15	80.0%	6.7%	0.0%	13.3%
2013	1	2	8	0	11	9.1%	18.2%	72.7%	0.0%
Total	22	6	27	8	63	34.9%	9.5%	42.9%	12.7%

The results above on type and access mode of nonpublic data are derived from a single journal's articles, and should be interpreted with caution. A more generalized assessment is difficult to undertake, since no journal in economics provides consistent data or metadata on the mode of access.

It is worth pointing out the increase in the past 2 decades of formal restricted-access data environments (RADEs), sponsored or funded by national statistical offices and funding agencies. RADE networks, with formal, nondiscriminatory, albeit often lengthy access protocols, have been set up in the United States (FSRDC) (D. H. Weinberg et al., 2007), Canada (Currie & Fortin, 2015), Germany (Bender & Heining, 2011), France (Bozio & Geoffard, 2017; Gadouche & Picard, 2017; "Une bulle pour protéger les fichiers," 2014), and many other countries. Often, these networks have been initiated by economists, though widespread use is made by other social scientists and in some cases health researchers. RADE are less common for private-sector data, although several initiatives have made progress and are frequently used by researchers: Institute for Research on Innovation and Science (IRIS, 2018; B. A. Weinberg et al., 2014), Health Care Cost Institute (HCCI, 2018; Newman et al., 2014), Private Capital Research Institute (PCRI) (Jeng & Lerner, 2016; PCRI, n.d.). When such nondiscriminatory agreements are implemented at scale, a significant number of researchers can obtain access to these data under strict security protocols. As of 2018, the FSRDC hosted more than 750 researchers on over 300 projects, of which 140 had started within the last 12 months (U.S. Census Bureau, 2019). The IAB FDZ lists over 500 projects active as of September 2019, most with multiple authors (Forschungsdatenzentrum des IAB, 2019). In these and other networks, many researchers share access to the same data sets, and could

potentially conduct reproducibility studies. Typically, access is via a network of secure rooms (FSRDC, Canada, Germany), but in some cases, remote access via 'thin clients' (France) or virtual desktop infrastructure (some Scandinavian countries, data from the Economic Research Service of the United States Department of Agriculture [USDA] via NORC) is allowed.

While RADE makes access easier, even reproducible, the complexity and cost of accessing these data generally remains higher than for public-use data. In theory, the need to manage and control access suggests that data curation processes at the data provider should be robust, these are typically not visible to end users. Requesting the same source data is often a challenge. The French and German RADE have recently started to assign DOI to restricted-access data sets (see, e.g., Schmucker et al., 2018). Support for researchers to create replication packages within RADEs is generally not well established (Lagoze & Vilhuber, 2017).

Some widely used data sets are accessible by any researcher, but the license they are subject to prevents their redistribution and thus their inclusion as part of data deposits. This includes nonconfidential data sets from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID) at the University of Michigan and data provided by IPUMS at the Minnesota Population Center. All of these data can be freely downloaded, subject to agreement to a license. IPUMS lists 963 publications for 2015 alone that use one of its data sources. The typical user will create a custom extract of the PSID and IPUMS databases through a data query system, not download specific data sets. Thus, each extract is essentially unique. Yet that same extract cannot be redistributed, or deposited at a journal or any other archive.[19] In 2018, the PSID, in collaboration with ICPSR, has addressed this issue with the PSID Repository, which allows researchers to deposit their custom extracts in full compliance with the PSID Conditions of Use.

Commercial ('proprietary') data is typically subject to licenses that also prohibit redistribution. Larger companies may have data provision as part of their service, but providing it to academic researchers is only a small part of the overall business. Dun and Bradstreet's Compustat, Bureau van Dijk's Orbis, Nielsen Scanner data via the Kilts Center at Chicago Booth (Kilts Center, n.d.), or Twitter data are all used frequently by economists and other social scientists. But providing robust and curated archives of data as used by clients over 5 or more years is typically not part of their service.[20] Typically, after signing an agreement, researchers can download the data, but remote access may also sometimes be used—the aforementioned PCRI provides access via the NORC Data Enclave. A novel method for unbiased and rules-based access to social media data has recently been proposed (King & Persily, 2018). In some cases, while researchers may be able to use the data at no cost, they may still be prevented from redistributing the data, as copyright is claimed on the database. A prominent example are stock indexes (S&P Dow Jones Indices LLC, 2020), which are highly visible to anybody with an internet connection or a newspaper, but subject to redistribution prohibitions. Most researchers also do not think to include or request redistribution rights in the acquisition contract, or at a minimum, the right to provide some level of access for the purpose of reproducibility. Nevertheless, such

agreements exist, but are often hard to find due to the opaque nature of the 'supplemental data' package on journal websites.[21]

4.4. The Importance of Transparent Public Data

While I point out the potential impact on reproducibility that restricted-access data may have, it is worthwhile to point out that even when data is shareable, there are issues related to reproducibility and replicability. While reproductions that identify errors in programs *used* by the researchers (Siskind, 1977; e.g., F. Welch, 1974, 1977) are testimony to the power of reproducible research, studies that have focused on errors in the production and appropriate use of public-use data are just as important. Widely used data sets, such as the Current Population Survey (CPS) and the American Community Survey (ACS), have a long history of use by academics, and have a vast amount of accompanying documentation. Thanks to methodology documentation, it has been possible to show that incorrect use of data can lead to misleading conclusions.[22] In some cases, previously undocumented errors in the data publication itself were discovered (Alexander et al., 2010). These are examples of replication studies, but also of the need for adequate documentation of the data collection, cleaning, and dissemination. Many other public-use data sets, as well as most researcher-collected data sets, lack the amount of documentation to support such transparency, critical for the downstream use of these data in research. For official statistics, the National Academies' workshop on "Transparency and Reproducibility in Federal Statistics" published a report with recommendations (NASEM, 2019).

4.5. Reproducible Research in Academic Education

One of the more difficult topics to empirically assess is the extent to which reproducibility is taught in economics, and to what extent in turn economic education is helped by reproducible data analyses. The extent of the use of replication exercises in economics classes is anecdotally high, but I am not aware of any study or survey demonstrating this. Most empirical economists teaching graduate economics classes will ask students to reproduce or replicate one or more relevant articles, though few of these replications are ever systematically made public if replications are successful,[23] though many failed reproductions and replications may have triggered articles and entire theses, attesting to the publication bias in replications. The most famous example in economics is, of course, the exchange between Reinhart and Rogoff, and graduate student Thomas Herndon, together with professors Pollin and Ash (Herndon et al., 2014; Reinhart & Rogoff, 2010).

It is worthwhile pointing out that the Canadian Research Data Center Network expedites access requests to confidential data for students, greatly facilitating work on master's and doctoral theses, and potentially opening the door to easier reproductions and replications using confidential data.

More recently, explicit training in reproducible methods (Ball & Medeiros, 2012; *Berkeley Initiative for Transparency in the Social Sciences*, 2015), and participation of economists in data science programs with reproducible methods has increased substantially, but again, no formal and systematic survey has been conducted.

5. Looking Ahead

Many of the issues facing reproducibility and replicability in economics are not unique to economics, and affect many other of the empirical social and clinical sciences. I touch here on a few of these topics.

5.1. Citing Data

A contributor to transparent and reproducible use of data is the ability to cite data, and to do so with precision. While data citation standards are well-established (Data Citation Synthesis Group & Martone, 2014; ICPSR, 2018; Starr et al., 2015), only recently have style guides at major economics journals provided a suggested data citation format. The Chicago or Harvard citation styles are generally followed, but as of the 15th edition, the *Chicago Manual of Style* does not provide strong guidance or examples on data citations ("Author-Date: Sample Citations," 2018). The *AER* now requires data sets to be cited, and provides a suggested data citation (*AER*, 2018; AEA, 2018) to supplement the Chicago style. Other journals, even when having an explicit reproducibility policy (Royal Economic Society, 2018), do not provide guidance on how to cite data sets. The absence of a data citation policy also affects incentives to create reproducible research (Galiani et al., 2017).

5.2. Big Data, Changing Data

Difficulties when citing data are compounded when the data is either changing, or is a potentially ill-defined subset of a larger static or dynamic databases. 'Big data' have always posed challenges—see the earlier discussion of the 1950s–1960s demand for access to government databases. By nature, they most often fall into the 'proprietary' and 'commercial' category, with the problems that entails for reproducibility. However, beyond the (solvable) problem of providing replicators with authorized access and enough computing resources to replicate original research, even defining or acquiring the original data inputs may be hard. Big data may be ephemerous by nature, too big to retain for significant duration (sometimes referred to as 'velocity'), temporally or cross-sectionally inconsistent (variable specifications change, sometimes referred to as 'variety'). This may make computational reproducibility impossible. However, replicability and generalizability studies can still be undertaken, as long as access to the same general data stream is repeatable. For instance, a study that uses data from an ephemerous social media platform where posts last no more than 24 hours ('velocity') and where the data schema may mutate over time ('variety') may not be computationally reproducible, because the posts will have been deleted (and terms of use may prohibit redistribution of any scraped data). But the same data collection (scraping or data extraction) can be repeated, albeit with some complexity in reprogramming to address the variety problem, leading to a replication study.

Changing data is not just an issue with what is commonly referred to as 'big data,' but also for more traditional very large data sets. IPUMS, as pointed out earlier, can only be accessed via a query interface. As of this writing, it is not possible through said interface to define the precise revision of the underlying database, the version of the query system used, and the exact query used. When the resulting extract cannot be redistributed, as is the case for the full population censuses, it is not feasible to reliably create reproducible analyses using

IPUMS, though the analysis could still be replicable. The same generic problem affects many other systems, though some data providers have found solutions. The PSID (*PSID - Data Center - Previous Carts*, 2018) allows for storage and later retrieval of the query parameters. The Census Bureau's OnTheMap (OnTheMap, 2018) provides a mechanism for users themselves to store reusable query parameters.

The above examples involve large, but slowly evolving databases. However, in the presence of big data, including data from social media (Facebook, Twitter), data and possibly data schemas evolve quite rapidly, and the simple mechanisms that the PSID, IPUMS, and the Census Bureau use, fail. Citation of such data sources remains an active research topic for institutions like the Research Data Alliance (2016; Rauber & Asmi, 2016), with no robust solution yet adopted.

While in theory, researchers are able to at least informally describe the data extraction and cleaning processes when run on third-party–controlled systems that are typical of big data, in practice, this does not happen. An informal analysis of various Twitter-related economics articles shows very little or no description of the data extraction and cleaning process. The problem, however, is not unique to big-data articles—most articles provide little if any input data cleaning code in reproducibility archives, in large part because provision of the code that manipulates the input data is only suggested, but not required by most data deposit policies.

5.3. Registration of Trials, Analysis Plans, and Reports

Related to concerns about replicability, but primarily aiming to address issues of publication bias and selective reporting of outcomes, the preregistration of research hypotheses, analysis plans, and trials has made inroads in economics, primarily in laboratory experiments and randomized control trials. Formal trial registries in the social sciences (including prominently in psychology) are inspired by similar efforts in the medical sciences (International Committee of Medical Journal Editors, 2020; National Libraries of Medicine, 2018). An early implementation in economics was the J-PAL Hypothesis Registry (The Abdul Latif Jameel Poverty Action Lab, 2009). In 2012, the *AEA* instantiated the AEA Randomized Control Trial (RCT) Registry (Katz et al., 2013; AEA Registry, 2020), "as a source of results for meta-analysis; as a one-stop resource to find out about available survey instruments and data."[24] The AEA RCT Registry keeps track of IRB protocol and approval numbers. Since 2017, registrations are reviewed for compliance with minimal criteria. As of May 2018, nearly 1,800 studies have been registered. Reference to the AEA RCT registry in published articles (of the AEA or elsewhere) has not been studied systematically. In 2019, the AEA RCT registry introduced DOIs for its registrations (K. Welch & Turitto, 2019).

Pre-analysis plans (PAP) offer similar benefits, without a particular focus on trials. Registries that allow to register both trials and PAPs include the Registry for International Development Impact Evaluations Registration (Dahl Rasmussen et al., 2011; RIDIE, n.d.), Evidence in Governance and Politics (2009), and AsPredicted (2018). The Open Science Framework (OSF) provides the ability to record snapshots of projects, providing a similar proof as formal registries (OSF, 2018a). Registration in general is voluntary, in contrast to

clinical trials (International Committee of Medical Journal Editors, 2020). Even without formal registries, several prominent articles have used PAPs to effectively frame their results, see (Christensen & Miguel, 2018) for some examples. Registrations and PAPs are not a panacea. While they may mitigate '*p*-hacking' (the manipulation or selective reporting of *p* values) (Brodeur et al., 2016), they may not increase the robustness of results (Coffman & Niederle, 2015). Some see registration generating disincentives for exploratory analysis (Coffman & Niederle, 2015), whereas others see exploratory analysis as a critical precursor to confirmatory, preregistered analyses (Nosek et al., 2018). Preregistration of complex and conditional hypotheses become quite difficult if not impossible to make (Olken, 2015). When researchers do preregister, they report spending 2 to 4 weeks preparing their materials (Ofosu & Posner, 2019). A key benefit may simply be to clarify which analyses were conceived without knowledge of the data, and which were developed post hoc, and to encourage "intellectual humility" (Nosek et al., 2019) by incentivizing researchers to plan their research ahead of time.

I note that several potential forms of PAPs are hiding in plain sight, precisely when early planning is not a choice, but a requirement. For instance, by making time-stamped research grant proposals or research data access requests (for RADEs) public, researchers could use such routinely submitted and, in the case of RADEs, compulsory documents as a form of PAPs (Lagoze & Vilhuber, 2017). Funders and data custodians could implement functionality within their submission systems to support such efforts, or explicitly encouraging researchers to routinely submit proposal documents at the relevant registries. Given the prevalence of restricted-access data sets, such a mechanism would have a potentially large and positive impact. To the best of my knowledge, this is not widely used at present.

An underappreciated tool that has most of the characteristics of PAPs is the use of validation and verification servers in combination with synthetic data (Kinney et al., 2011; Reiter, 2003; Reiter et al., 2009; Vilhuber et al., 2016). When using synthetic data, researchers build sophisticated models using data that is not guaranteed to provide the correct inferences (Vilhuber & Abowd, 2016). By submitting their code for validation, researchers are in effect submitting a PAP. Various U.S. statistical agencies as well as those in other countries (Drechsler, 2012; Nowok et al., 2016) have been experimenting with these methods, though primarily as a means to alleviate the access barriers, not as a tool to address reproducibility.

Registered reports carry the idea of preregistration further, and condition the publication of an article only on the prespecified analysis. Not only do the authors have no (significant) leeway in analyzing the data, but the editors and reviewers also cannot select publications based on the statistical results (Chambers, 2014; Nosek & Lakens, 2014; OSF, 2018b). Registered reports are intended to counter the publication bias in favor of 'significant' results, and encourage replications regardless of outcomes. An overview of the current state of registered reports is provided in Hardwicke & Ioannidis (2018). As of 2018, registered reports are uncommon in economics.[25]

A concern occasionally heard is that preregistration, submission of proposals to funders and data gatekeepers, and registered reports require authors to put promising ideas prematurely into the public domain. While the risk

of unethical appropriation of ideas by those with a privileged view onto the early submissions can never be entirely discarded, that problem is not restricted to the mechanisms mentioned here.[26] The risk is mitigated to some extent by the fact that most mechanisms here either allow for or require an embargo period for the registered information.[27]

5.4. Published Replications

Registered reports are seen as a potential solution to obtain more published reproducibility studies (Galiani et al., 2017). Because most reproducibility studies of individual articles 'only' confirm existing results, they fail the 'novelty test' that most editors apply to submitted articles (Galiani et al., 2017). Berry and coauthors (2017) analyzed all papers in Volume 100 of the *AER*, identifying how many were referenced as part of replication or cited in follow-on work. While partially confirming earlier findings that strongly cited articles will also be replicated (Hamermesh, 2007), the authors found that 60% of the original articles were referenced in replication or extension work, but only 20% appeared in explicit replications. Of the roughly 1,500 papers that cite the papers in the volume, only about 50 (3.5%) are replications, and of those, only 8 (0.5%) focused explicitly on replicating one paper. Out of roughly 2,600 articles in the *AER* between 2004 and 2016, the ReplicationWiki (Höffler, 2017a) identifies 44 'Comments' as 'replications' of some sort. A few journals have introduced specific sections for reproducibility studies, following the longtime lead of the *Journal of Applied Econometrics*. Some journals have had calls for special issues dedicated to specific replication studies (Burman et al., 2010).

Even rarer are studies that conduct replications prior to their publication, of their own volition. Antenucci et al. (2014) predict the unemployment rate from Twitter data. After having written the paper, they continued to update the statistics on their website ("Prediction of Initial Claims for Unemployment Insurance," 2017), thus effectively replicating their paper's results on an ongoing basis. Shortly after release of the working paper, the model started to fail. The authors posted a warning on their website in 2015, but continued to publish new data and predictions until 2017, in effect, demonstrating themselves that the originally published model did not generalize. Similarly, Bowers et al. (2017) present their original experiment, and their own failure to replicate the original results when conducting the experiment a second time.

5.5. Elevating the Importance of Data and Code Availability

Enabling easier publication of replication studies and reproductions is one approach that will likely enhance overall reproducibility of economic research. A complementary approach is to make the analysis of code and data associated with the research a part of the peer review process: a stronger emphasis on prepublication verification of data and code packages (Jacoby et al., 2017), and the timely availability of results of such tests to referees and editors, prior to final decisions on acceptance. The author of this article was appointed as data editor (Duflo & Hoynes, 2018) with the task to review not just the data availability policy, but also the methods and procedures supporting the implementation of the policy. I have outlined a vision for increased transparency

and reproducibility of the materials supporting articles published in the association's journals (Vilhuber, 2019), and announced a new policy with stronger enforcement, prepublication verification of data deposits and computer code (AEA, 2019). These concerns are not unique to the AEA, with similar considerations and activity under way at the *Review of Economic Studies*, the *Economic Journal*, and the *Econometric Journal* of the Royal Society.

At statistical agencies and RADEs, reproducibility and its interaction with access restrictions and the protection of confidentiality is being discussed. At the U.S. Census Bureau and the Canadian Research Data Centers, working groups are looking into how the visibility of reproducibility within secure research environments can be increased.[28] Most of the processes in place to ensure confidentiality actually imply reproducibility, since research results and methods are vetted by reviewers before being released to the public. It should thus be relatively straightforward to demonstrate reproducibility of studies that are conducted in these environments (Lagoze & Vilhuber, 2017). In 2019, *cascad* (*Certification Agency for Scientific Code & Data*, n.d.) was launched as the 'first' service to provide reproducibility services (Pérignon et al., 2019) both within the French research data system *Centre d'accès sécurisé aux données* (CASD - *Centre d'accès Sécurisé Aux Données*, n.d.) and for manuscripts using public use data.

Multiple research *institutions* are not waiting for journals to implement more stringent criteria. For instance, projects at J-PAL that collect data with funding from their research initiatives are subject to a data availability policy (The Abdul Latif Jameel Poverty Action Lab, 2015), and all J-PAL affiliated researchers are encouraged to publish their data sets in a Dataverse. Some research institutions offer a 'code check' service to researchers prior to submission to journal (CISER, 2018). The Federal Reserve Bank of Kansas City is implementing properly curated data supplements for its working papers, prior to any journal publication (Butler & Kulp, 2018).

6. Conclusion

Reproducibility has certainly gained more visibility and traction since Dewald *et al.'s* (1986) wake-up call. Twenty years after Dewald *et al.* saw the emergence of data archives and data availability policies at top economics journals. Thirty years after Dewald *et al.*, the largest association of economists has designated a data editor for its journals, and is implementing pervasive prepublication reproducibility checks, including in certain cases when data cannot be published. More general projects that provide training on reproducibility (TIER, Teaching Integrity in Empirical Research, Ball & Medeiros, 2012; BITSS, Berkeley Initiative for Transparency in the Social Sciences, 2015) and infrastructure for curated reproducibility (CodeOcean, 2018; Dataverse, n.d.; OpenICPSR, n.d.; RunMyCode, Stodden et al., 2012; Whole Tale, Brinckman et al., 2018; Zenodo, n.d.) are gaining traction in economics, and many other initiatives that are likely to yield improved reproducibility are in their early stages.

Still, after 30 years, the results of reproducibility studies consistently show problems with about a third of reproduction attempts, and the increasing share of restricted-access data in economic research requires new tools, procedures, and methods to enable greater visibility into the reproducibility of such studies. Incorporating consistent training in reproducibility into graduate curricula remains one of the challenges for the (near) future.

As shown in this article, economics has a long history of open sharing of prepublication manuscripts (as working papers) and code (for instance, through the Statistical Software Components [SSC]). Even though there are no strong indications (yet) of a tendency to use open-source software—sometimes put forward as a contributor to greater reproducibility—the pervasive use of commercial closed-source software has not seemed to materially impact the computational reproducibility of research. The need to provide replication materials to comply with data policies as part of the scholarly publication process emerged organically nearly 20 years ago. Compliance at most reputable journals has been high, and has been endogenized by researchers.

None of these observations are unique to economics. Viewed in combination, they suggest that newer, stronger methods—prepublication reproducibility checks, registered reports, methods to alleviate *p*-hacking—may be feasibly implemented in economics, and if successful, can inform other disciplines on a similar path.

Acknowledgments

Disclosure Statement

References

The Abdul Latif Jameel Poverty Action Lab. (2009). *Hypothesis registry.*
https://www.povertyactionlab.org/Hypothesis-Registry

The Abdul Latif Jameel Poverty Action Lab. (2015). *Transparency & reproducibility*.
https://www.povertyactionlab.org/research-resources/transparency-and-reproducibility

Alexander, J. T., Davern, M., & Stevenson, B. (2010). Inaccurate age and sex data in the census PUMS files: Evidence and implications. *Public Opinion Quarterly, 74*(3), 551–569. https://doi.org/10.1093/poq/nfq033

American Economic Association. (2019, July 16). *AEA member announcements: Updated AEA data and code availability policy*. https://www.aeaweb.org/news/member-announcements-july-16-2019

American Economic Association. (2018). *Sample references—Styles of the AEA*.
https://www.aeaweb.org/journals/policies/sample-references

American Economic Association RCT Registry. (2020). *FAQ*. https://www.socialscienceregistry.org/site/faq

American Economic Review. (2018). *AER style guide for accepted articles*.
https://www.aeaweb.org/journals/aer/submissions/accepted-articles/styleguide

Anderson, M. J. (2015). *The American census: A social history* (2d ed.). Yale University Press.

Anderson, R. G., & Dewald, W. G. (1994). Replication and scientific standards in applied economics a decade after the *Journal of Money, Credit and Banking* Project. *Federal Reserve Bank of St. Louis Review, 76*(6), 79–83. https://doi.org/10.20955/r.76.79-83

Antenucci, D., Cafarella, M., Levenstein, M., Ré, C., & Shapiro, M. (2014). *Using social media to measure labor market flows*. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w20010.

Arguillas, F., Christian, T.-M., & Peer, L. (2018). *Education for (a) CURE: Developing a prescription for training in data curation for reproducibility*. IASSIST.

ASA Advisory Committee. (1959). Recommendations on availability of federal statistical materials to nongovernmental research workers. *The American Statistician, 13*(4), 15–37.
http://www.jstor.org/stable/2685752

AsPredicted. (2018). *About*. https://aspredicted.org/messages/about.php

Author-Date: Sample Citations. (2018). *The Chicago Manual of Style online*.
http://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-2.html

Ball, R., & Medeiros, N. (2012). Teaching integrity in empirical research: A protocol for documenting data management and analysis. *The Journal of Economic Education, 43*(2), 182–189.
https://doi.org/10.1080/00220485.2012.659647

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2013a). *Replication data for: The nature of risk preferences: Evidence from Insurance Choices* (Version 1) [Data set]. American Economic Association [publisher]. ICPSR – Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E116116V1

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2013b). The nature of risk preferences: Evidence from insurance choices. *American Economic Review, 103*(6), 2499–2529. https://doi.org/10.1257/aer.103.6.2499

Bátiz-Lazo, B., & Krichel, T. (2012). A brief business history of an on-line distribution system for academic research called NEP, 1998-2010. *Journal of Management History, 18*(4), 445–468. https://doi.org/10.1108/17511341211258765150788

Bender, S., & Heining, J. (2011). The Research-Data-Centre in Research-Data-Centre approach: A first step towards decentralised international data sharing. *IASSIST Quarterly / International Association for Social Science Information Service and Technology, 35*(3). https://doi.org/10.29173/iq119

Berkeley Initiative for Transparency in the Social Sciences. (2015, October 8). *About*. https://www.bitss.org/about/

Berkeley Initiative for Transparency in the Social Sciences. (2018). *About registered reports at the JDE*. https://www.bitss.org/resources/registered-reports-at-the-journal-of-development-economics/

Berry, J., Coffman, L. C., Hanley, D., Gihleb, R., & Wilson, A. J. (2017). Assessing the rate of replication in economics. *American Economic Review, 107*(5), 27–31. https://doi.org/10.1257/aer.p20171119

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review, 59*(1), 65–98. https://doi.org/10.1137/141000671

BibEc. (1997, December 11). *Main page*. http://web.archive.org/web/19971211044921/http://netec.mcc.ac.uk:80/BibEc.html

Bollen, K., Cacioppo, J. T., Kaplan, R. M., Krosnick, J. A., & Olds, J. L. (2015). *Social, behavioral, and economic sciences perspectives on robust and reliable science* [Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences]. National Science Foundation. https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

Bowers, J., Higgins, N., Karlan, D., Tulman, S., & Zinman, J. (2017). Challenges to replication and iteration in field experiments: Evidence from two direct mail shots. *American Economic Review, 107*(5), 462–465. https://doi.org/10.1257/aer.p20171060

Bozio, A., & Geoffard, P.-Y. (2017). L'accès des chercheurs aux données administratives [Researcher access to administrative data] (p. 68) [Rapport au secretaire d'état charge de l'industrie, du numerique et de l'innovation]. Conseil national de l'information statistique. https://www.economie.gouv.fr/files/files/PDF/2017/Rapport_CNIS_04_2017.pdf

Brinckman, A., Chard, K., Gaffney, N., Hategan, M., Jones, M. B., Kowalik, K., Kulasekaran, S., Ludäscher, B., Mecum, B. D., Nabrzyski, J., Stodden, V., Taylor, I. J., Turk, M. J., & Turner, K. (2018). Computing environments for reproducibility: Capturing the "Whole Tale." *Future Generation Computer Systems*, *94*, 854–867. https://doi.org/10.1016/j.future.2017.12.029

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics*, *8*(1), 1–32. https://doi.org/10.1257/app.20150044

Bullock, C. J. (1919). Prefatory statement. *The Review of Economics and Statistics*, *1*(1). http://www.jstor.org/stable/1928753

Burman, L. E., Reed, W. R., & Alm, J. (2010). A call for replication studies. *Public Finance Review*, *38*(6), 787–793. https://doi.org/10.1177/1091142110385210

Butler, C., & Kulp, C. (2018). *The role of data supplements in reproducibility: Curation challenges* [Conference presentation]. The 2018 Annual International Association for Social Science Information Services & Technology (IASSIST) Conference, Montreal, Quebec, Canada

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 433–1436. https://doi.org/10.1126/science.aaf0918

cascad - Certification Agency for Scientific Code & Data. (n.d.). Retrieved September 22, 2019, from https://www.cascad.tech/

CASD - Centre d'accès sécurisé aux données. (n.d.). Retrieved September 22, 2019, from https://www.casd.eu/

Chambers, C. (2014). *Registered reports: A step change in scientific publishing. Reviewers' update*. Elsevier. https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/registered-reports-a-step-change-in-scientific-publishing

Chang, A. C., & Li, P. (2017). A preanalysis plan to replicate sixty economics research papers that worked half of the time. *American Economic Review*, *107*(5), 60–64. https://doi.org/10.1257/aer.p20171034

Chemin, M., & Wasmer, E. (2009). Using Alsace-Moselle local laws to build a difference-in-differences estimation strategy of the employment effects of the 35-hour workweek regulation in France. *Journal of Labor Economics, 27*(4), 487–524. https://doi.org/10.1086/605426

Chemin, M., & Wasmer, E. (2017). Erratum. *Journal of Labor Economics, 35*(4), 1149–1152. https://doi.org/10.1086/693983

Chetty, R. (2012). *Time trends in the use of administrative data for empirical research*. NBER Summer Institute. http://www.rajchetty.com/chettyfiles/admin_data_trends.pdf

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature, 56*(3), 920–980. https://doi.org/10.1257/jel.20171350

Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys, 31*(1), 326–342. https://doi.org/10.1111/joes.12139

Code Ocean. (2018). *About Code Ocean*. https://codeocean.com/about

CodEc. (1998, January 21). *Programs for economics and econometrics.* http://web.archive.org/web/19980121224535/http://netec.mcc.ac.uk:80/CodEc.html

Coffman, L. C., & Niederle, M. (2015). Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives, 29*(3), 81–98. https://doi.org/10.1257/jep.29.3.81

Cornell Institute for Social and Economic Research. (2018). *Results reproduction (R-squared)*. https://ciser.cornell.edu/research/results-reproduction-r-squared-service/

Cox, N. J. (2010). A conversation with Kit Baum. *The Stata Journal, 10*(1), 3–8. https://doi.org/10.1177/1536867X1001000102

Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018). Data policies of highly-ranked social science journals. *SocArXiv*. https://doi.org/10.17605/osf.io/9h7ay

Cruz, J. M. B., & Krichel, T. (2000). Cataloging economics preprints. *Journal of Internet Cataloging, 3*(2–3), 227–241. https://doi.org/10.1300/J141v03n02_08

Currie, R., & Fortin, S. (2015). *Social statistics matter: History of the Canadian Research Data Center Network*. Canadian Research Data Centre Network. http://rdc-cdr.ca/sites/default/files/social-statistics-matter-crdcn-history.pdf

Dahl Rasmussen, O., Malchow-Møller, N., & Barnebeck Andersen, T. (2011). Walking the talk: The need for a trial registry for development interventions. *Journal of Development Effectiveness, 3*(4), 502–519.

https://doi.org/10.1080/19439342.2011.605160

Data Citation Synthesis Group, & Martone, M. (2014). *Joint declaration of data citation principles*. Force11. https://doi.org/10.25490/a97f-egyk

Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in empirical economics: The *Journal of Money, Credit and Banking* Project. *The American Economic Review, 76*(4), 587–603. https://www.jstor.org/stable/1806061

Drechsler, J. (2012). New data dissemination approaches in old Europe – synthetic datasets for a German establishment survey. *Journal of Applied Statistics, 39*(2), 243–265. https://doi.org/10.1080/02664763.2011.584523

Duflo, E. (2018). Report of the editor: *American Economic Review*. *AEA Papers and Proceedings, 108*, 636–651. https://doi.org/10.1257/pandp.108.636

Duflo, E., & Hoynes, H. (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

Duvendack, M., Palmer-Jones, R., & Reed, W. R. (2017). What is meant by "replication" and why does it encounter resistance in economics? *American Economic Review, 107*(5), 46–51. https://doi.org/10.1257/aer.p20171031

Duvendack, M., Palmer-Jones, R. W., & Reed, W. (2015). Replications in economics: A progress report. *Econ Journal Watch, 12*(2), 164–191.

Economics Software | IDEAS/RePEc. (n.d.). Retrieved July 19, 2018, from https://ideas.repec.org/i/c.html

Eddelbüttel, D. (1997). A code archive for economics and econometrics. *Computational Economics, 10*(4). http://web.archive.org/web/19980515055648/http://netec.mcc.ac.uk:80/~adnetec/CodEc/ce97.pdf

Evidence in Governance and Politics. (2009). *Rules and procedures*. Retreived July 22, 2018, from https://web.archive.org/web/20191124223125/http://egap.org/sites/default/files/pdfs/20110608_EGAP_structure.pdf

Feige, E. (1975). The consequences of journal editorial policies and a suggestion for revision. *Journal of Political Economy, 83*(6), 1291–1295. http://doi.org/10.1086/260400

Forschungsdatenzentrum des IAB. (2019, September 20). *Projects at FDZ.* http://doku.iab.de/fdz/projekte/Nutzer_Projekte_Liste.pdf

Frisch, R. (1933). Editor's note. *Econometrica, 1*(1), 1–4.

Gadouche, K., & Picard, N. (2017). L'accès aux données très détaillées pour la recherche scientifique [Access to detailed data for scientific research] (Working Paper No. 2017–06; THEMA, p. 18). Université Cergy-Pontoise. https://www.casd.eu/wp/wp-content/uploads/L_acces_aux_donnees_tres_detaillees_pour_la_recherche_scientifique.pdf

Galiani, S., Gertler, P., & Romero, M. (2017). *Incentives for replication in economics* (Working Paper No. 23576). National Bureau of Economic Research. https://doi.org/10.3386/w23576

Ginsparg, P. (1997). Winners and losers in the global research village. *The Serials Librarian, 30*(3–4), 83–95. https://doi.org/10.1300/J123v30n03_13

Glandon, P. (2011). *Report on the American Economic Review Data Availability Compliance Project. Appendix to American Economic Review Editors Report*. American Economic Association. http://www.aeaweb.org/aer/2011_Data_Compliance_Report.pdf

Godechot, O. (2016a). *Can we use Alsace-Moselle for estimating the employment effects of the 35-hour workweek regulation in France?* [Mimeo]. SciencesPo - Observatoire Sociologique du Changement. http://olivier.godechot.free.fr/hopfichiers/fichierspub/Comment_on_Chemin_Wasmer_2009_Jole.pdf

Godechot, O. (2016b). *L'Alsace-Moselle peut-elle décider des 35 heures? [Can Alsace-Moselle decide on the 35h work schedule?]*(Notes et Documents de l'OSC ND 2016-04). SciencesPo - Observatoire Sociologique du Changement. http://www.sciencespo.fr/osc/sites/sciencespo.fr.osc/files/ND_2016-04.pdf

Goldberg, P. K. (2017). Report of the editor: *American Economic Review*. *American Economic Review, 107*(5), 699–712. https://doi.org/10.1257/aer.107.5.699

Halpern, J. Y. (1998). A computing research repository. *D-Lib Magazine, 4*(11). https://doi.org/10.1045/november98-halpern

Hamermesh, D. S. (2007). Viewpoint: Replication in economics. *Canadian Journal of Economics, 40*(3), 715–733. https://doi.org/10.1111/j.1365-2966.2007.00428.x

Hamermesh, D. S. (2017). Replication in labor economics: Evidence from data, and what it suggests. *American Economic Review, 107*(5), 37–40. https://doi.org/10.1257/aer.p20171121

Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Mapping the universe of registered reports. *Nature Human Behaviour, 2*(11), 793–796. https://doi.org/10.1038/s41562-018-0444-y

Health Care Cost Institute. (2018). *About the Health Care Cost Institute*. http://www.healthcostinstitute.org/about-hcci/

Herndon, T., Ash, M., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics, 38*(2), 257–279. https://doi.org/10.1093/cje/bet075

Hirsch, B. T., & Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics, 22*(3), 689–722. https://doi.org/10.1086/383112

Höffler, J. H. (2017a). ReplicationWiki: Improving transparency in social sciences research. *D-Lib Magazine, 23*(3/4). https://doi.org/10.1045/march2017-hoeffler

Höffler, J. H. (2017b). Replication and economics journal policies. *American Economic Review, 107*(5), 52–55. https://doi.org/10.1257/aer.p20171032

Inter-University Consortium of Political and Social Research. (n.d.). *ICPSR: The founding and early years*. Retrieved July 18, 2018, from https://www.icpsr.umich.edu/web/pages/about/history/early-years.html

Inter-University Consortium of Political and Social Research. (2018). *Data citations*. https://web.archive.org/web/20180616123222/https://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/citations.jsp

IDEAS/RePEc. (n.d.). Retrieved July 19, 2018, from https://ideas.repec.org/

International Committee of Medical Journal Editors. (2020). FAQ: Clinical trials registration. http://www.icmje.org/about-icmje/faqs/clinical-trials-registration/

Institute for Research on Innovation and Science. (2018). *About*. http://iris.isr.umich.edu/about/

Jacoby, W. G., Lafferty-Hess, S., & Christian, T.-M. (2017, July 17). Should journals be responsible for reproducibility? *Inside Higher Ed*. https://www.insidehighered.com/blogs/rethinking-research/should-journals-be-responsible-reproducibility

Jeng, L., & Lerner, J. (2016). Making private data accessible in an opaque industry: The experience of the Private Capital Research Institute. *American Economic Review, 106*(5), 157–160. https://doi.org/10.1257/aer.p20161059

Journal of Applied Econometrics. (2014). Extension of the replication section's coverage. https://onlinelibrary.wiley.com/page/journal/10991255/homepage/News.html#replication

Journal of Development Economics. (2019, July 17). Registered reports at JDE: Lessons learned so far. https://www.journals.elsevier.com/journal-of-development-economics/announcements/registered-reports-at-jde

Katz, L., Duflo, E., Goldberg, P., & Thomas, D. (2013, November 28). *Email: AEA registry for controlled trials*. American Economic Association. https://web.archive.org/web/20131128040053/http://www.aeaweb.org:80/announcements/20131118_rct_email.php

Kilts Center. (n.d.). *Marketing datasets*. Retrieved June 14, 2020, from https://www.chicagobooth.edu/research/kilts/datasets

King, G. (1995). Replication, replication. *PS, Political Science & Politics, 28*(3), 443–499. https://tinyurl.com/mvc5kg5

King, G., & Persily, N. (2018). *A new model for industry-academic partnerships* [Working Paper]. Harvard University. http://j.mp/2q1IQpH

Kingi, H., Stanchi, F., Vilhuber, L., & Herbert, S. (2018). *The reproducibility of economics research: A case study.* BITSS Annual Meeting, Berkeley, CA, December 10 [Presentation]. Open Science Framework. https://osf.io/srg57/

Kinney, S. K., Reiter, J. P., Reznek, A. P., Miranda, J., Jarmin, R. S., & Abowd, J. M. (2011). Towards unrestricted public use business microdata: The Synthetic Longitudinal Business Database. *International Statistical Review, 79*(3), 362–384. https://doi.org/10.1111/j.1751-5823.2011.00153.x

Kraus, R. (2013). Statistical déjà vu: The National Data Center Proposal of 1965 and its descendants. *Journal of Privacy and Confidentiality, 5*(1). https://doi.org/10.29012/jpc.v5i1.624

Krichel, T. (1997). WoPEc: Electronic Working Papers in Economics Services. *Ariadne*, (8). http://www.ariadne.ac.uk/issue8/wopec

Krichel, T., & Zimmermann, C. (2009). The economics of open bibliographic data provision. *Economic Analysis and Policy, 39*(1), 143–152. https://doi.org/10.1016/S0313-5926(09)50049-5

Lagoze, C., & Vilhuber, L. (2017). Making confidential data part of reproducible research. *Chance*. http://chance.amstat.org/2017/09/reproducible-research/

Margo, R. A. (2011). The economic history of the *American Economic Review*: A century's explosion of economics research. *American Economic Review, 101*(1), 9–35. https://doi.org/10.1257/aer.101.1.9

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the *JMCB* archive. *Journal of Money, Credit, and Banking, 38*(4), 1093–1107. https://doi.org/10.1353/mcb.2006.0061

Miller, W. E. (1963). The Inter-University Consortium for Political Research. *American Behavioral Scientist, 7*(3), 11. https://doi.org/10.1177/000276426300700304

Moffitt, R. A. (2011). Report of the editor: *American Economic Review* (with Appendix by Philip J. Glandon). *American Economic Review, 101*(3), 684–693. https://doi.org/10.1257/aer.101.3.684

National Academies of Sciences, Engineering, and Medicine. (2019). *Methods to Foster Transparency and Reproducibility of Federal Statistics*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25305

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. National Academies Press. https://doi.org/10.17226/25303

National Libraries of Medicine. (2018). *ClinicalTrials.gov fact sheet*. http://wayback.archive-it.org/org-350/20180312141720/https://www.nlm.nih.gov/pubs/factsheets/clintrial.html

National Research Council. (2003). *Sharing publication-related data and materials: Responsibilities of authorship in the life sciences*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10613

Newman, D., Herrera, C.-N., & Parente, S. T. (2014). Overcoming barriers to a research-ready national commercial claims database. *American Journal of Managed Care, 11*(17), eSP25–eSP30. https://www.ajmc.com/journals/issue/2014/2014-11-vol20-sp/overcoming-barriers-to-a-research-ready-national-commercial-claims-database

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., Veer, A. E. van 't, & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences, 23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

Nowok, B., Raab, G., & Dibben, C. (2016). Synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software, 74*(11), 1–26. https://doi.org/10.18637/jss.v074.i11

Ofosu, G., & Posner, D. N. (2019). *Pre-analysis plans: A stocktaking* (MetaArXiv Preprints No. e4pum). Berkeley Initiative for Transparency in the Social Sciences (BITSS). https://doi.org/10.31222/osf.io/e4pum

Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives, 29*(3), 61–80. https://doi.org/10.1257/jep.29.3.61

OnTheMap. (2018). https://onthemap.ces.census.gov/

OpenICPSR. (n.d.). *Share your behavioral health and social science research data*. Retrieved July 22, 2018, from https://www.openicpsr.org/openicpsr/

Open Science Framework. (2018a). *Guides—Registrations*. http://help.osf.io/m/registrations

Open Science Framework. (2018b). *Registered reports*. https://cos.io/rr/

Panel Study of Income Dynamics. (2018). *About the PSID* repository. OpenICPSR. https://www.openicpsr.org/openicpsr/psid

The Private Capital Research Institute. (n.d.). *About.* Retrieved July 20, 2018, from http://www.privatecapitalresearchinstitute.org/about.php

Pérignon, C., Gadouche, K., Hurlin, C., Silberman, R., & Debonnel, E. (2019). Certify reproducibility with confidential data. *Science, 365*(6449), 127–128. https://doi.org/10.1126/science.aaw2825

Pesaran, H. (2003). Introducing a replication section. *Journal of Applied Econometrics, 18*(1), 111. https://doi.org/10.1002/jae.709

Prediction of initial claims for unemployment insurance. (2017). Retreived July 22, 2018, from http://econprediction.eecs.umich.edu/

Panel Study of Income Dynamics. (2018). *Data Center—Previous carts*. University of Michigan.

R Core Team. (2000). *R: A Language and Environment for Statistical Computing* (Version 1.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rauber, A., & Asmi, A. (2016). Identification of reproducible subsets for data citation, sharing and re-use. *Bulletin of IEEE Technical Committee on Digital Libraries, 12*(1), 10. https://bulletin.jcdl.org/Bulletin/v12n1/papers/IEEE-TCDL-DC-2016_paper_1.pdf

Registry for International Development Impact Evaluations. (n.d.). Retrieved July 22, 2018, from http://ridie.org

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a time of debt. *American Economic Review, 100*(2), 573–578. https://doi.org/10.1257/aer.100.2.573

Reiter, J P. (2003). Model diagnostics for remote-access regression servers. *Statistics and Computing, 13*(4), 371–380. https://doi.org/10.1023/A:1025623108012

Reiter, J. P, Oganian, A., & Karr, A. F. (2009). Verification servers: Enabling analysts to assess the quality of inferences from public use data. *Computational Statistics & Data Analysis, 53*(4), 1475–1482. https://doi.org/10.1016/j.csda.2008.10.006

RePEc: Research Papers in Economics. (n.d.). Retrieved July 19, 2018, from http://repec.org/

Research Data Alliance. (2016, November 4). *Data versioning WG*. https://www.rd-alliance.org/groups/data-versioning-wg

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise*. Elsevier. https://doi.org/10.1016/C2017-0-01828-1

Royal Economic Society. (2018). *Submissions—Royal Economic Society*. http://www.res.org.uk/view/submissionsEconometrics.html

Sargent, T. J., & Stachurski, J. (2017). *Lectures in quantitative economics*. Quantitative Economics. https://lectures.quantecon.org/about_lectures.html

Schmucker, A., Ganzer, A., Stegmaier, J., & Wolter, S. (2018). Betriebs-Historik-Panel 1975-2017. [Establishment History Panel 1975-2017]. Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB). https://doi.org/10.5164/IAB.BHP7517.DE.EN.V1

Schöpfel, J. (2011). Towards a Prague definition of grey literature. *The Grey Journal, 7*(1). http://greynet.org/images/GL12_S1P,_Sch_pfel.pdf

Siskind, F. B. (1977). Minimum wage legislation in the United States: Comment. *Economic Inquiry, 15*(1), 135–138. https://doi.org/10.1111/j.1465-7295.1977.tb00457.x

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations* (2007 edition, S. M. Soares). MetaLibri Digital Library.

Sobek, M., & Ruggles, S. (1999). The IPUMS Project: An update. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 32*(3), 102–110. https://doi.org/10.1080/01615449909598930

S&P Dow Jones Indices LLC. (2020). *S&P 500* [SP500]. FRED, Federal Reserve Bank of St. Louis [distributor]. https://fred.stlouisfed.org/series/SP500

Standard charts and tables: Original data. (1919). *The Review of Economics and Statistics, 1*(1), 64–103. https://doi.org/10.2307/1928764

Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R. R., Duerr, R., Haak, L. L., Haendel, M., Herman, I., Hodson, S., Hourclé, J., Kratz, J. E., Lin, J., Nielsen, L. H., Nurnberger, A., Proell, S., Rauber, A., Sacchi, S., Smith, A., … Clark, T. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science, 1*, Article e1. https://doi.org/10.7717/peerj-cs.1

Stigler, G. J., Stigler, S. M., & Friedland, C. (1995). The journals of economics. *Journal of Political Economy, 103*(2), 331–359. http://doi.org/10.1086/261986

Stodden, V., Guo, P., & Ma, Z. (2013). Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLoS ONE, 8*(6), Article e67111. https://doi.org/10.1371/journal.pone.0067111

Stodden, V., Hurlin, C., & Perignon, C. (2012). RunMyCode.Org: A novel dissemination and collaboration platform for executing published computational results. *SSRN*. https://doi.org/10.2139/ssrn.2147710

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences, 115*(11), Article 201708290. https://doi.org/10.1073/pnas.1708290115

Dataverse. (n.d.). *The Dataverse Project*. Retrieved July 19, 2018, from https://dataverse.org/home

Une bulle pour protéger les fichiers. (2014, April 9). *Le Monde*, 4–5. https://www.casd.eu/wp/wp-content/uploads/2017/06/2014-04-09-Le-Monde-Science-M%C3%A9decine.pdf

U.S. Census Bureau. (2019). *Center for Economic Studies and Research Data Centers Research report: 2018*. U.S. Census Bureau Center for Economic Studies. https://www.census.gov/ces/pdf/2018_CES_Annual_Report.pdf

Vilhuber, L. (2019). Report by the AEA data editor. *AEA Papers and Proceedings, 109*, 718–729. https://doi.org/10.1257/pandp.109.718

Vilhuber, L. (2020). *Data for: Requesting replication materials via email*. Labor Dynamics Institute. https://doi.org/10.5281/ZENODO.4267155

Vilhuber, L., & Abowd, J. (2016, May 6). *Usage and outcomes of the Synthetic Data Server* [Presentation]. Annual Meeting of the Society of Labor Economists, Seattle, WA. http://hdl.handle.net/1813/43883

Vilhuber, L., Abowd, J. M., & Reiter, J. P. (2016). Synthetic establishment microdata around the world. *Statistical Journal of the International Association for Official Statistics, 32*(1), 65–68. https://doi.org/10.3233/SJI-160964

Vilhuber, L., Turitto, J., & Welch, K. (2020). Report by the AEA data editor. *AEA Papers and Proceedings, 110*, 764–775. https://doi.org/10.1257/pandp.110.764

Vlaeminck, S., & Herrmann, L.-K. (2015). Data policies and data archives: A new paradigm for academic publishing in economic sciences? In B. Schmidt, & M. Dobreva (Eds.), *Proceedings of the 19th International Conference on Electronic Publishing* (pp. 145–155). IOS Press. https://doi.org/10.3233/978-1-61499-562-3-145

Weinberg, B. A., Owen-Smith, J., Rosen, R. F., Schwarz, L., Allen, B. M., Weiss, R. E., & Lane, J. (2014). Science funding and short-term economic activity. *Science, 344*(6179), 41–43. https://doi.org/10.1126/science.1250055

Weinberg, D. H., Abowd, J. M., Steel, P. M., Zayatz, L., & Rowland, S. K. (2007). *Access methods for United States microdata* (No. 07–25). Center for Economic Studies, U.S. Census Bureau. https://doi.org/10.2139/ssrn.1015374

Welch, F. (1973). *Education, information, and efficiency* (No. w0001). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w0001/w0001.pdf

Welch, F. (1974). Minimum wage legislation in the United States. *Economic Inquiry, 12*(3), 285–318. https://doi.org/10.1111/j.1465-7295.1974.tb00401.x

Welch, F. (1977). Minimum wage legislation in the United States: Reply. *Economic Inquiry, 15*(1), 139–142. https://doi.org/10.1111/j.1465-7295.1977.tb00458.x

Welch, K., & Turitto, J. (2019, August 15). Improving research transparency through easier, faster access to studies in the AEA RCT Registry. *J-PAL*. https://www.povertyactionlab.org/blog/8-15-19/improving-research-transparency-through-easier-faster-access-studies-aea-rct-registry

Zenodo. (n.d.). *Research. Shared*. Retrieved July 22, 2018, from http://about.zenodo.org/

Appendix

Table A1. Data in publications—1960.

Publication	Volume	Number of Articles	Empirical Articles	Public-Use Only	Percent Empirical	Percent Public
AER	50	21	13	10	61.9%	76.9%
Econometrica	28	46	15	8	32.6%	53.3%
ReStat	42	33	23	15	69.7%	65.2%

Methodology: We identified all articles in the relevant volumes (all 1960). An article was empirical if it used some data set, and was 'public-use only' if all data sets used could be identified as publicly accessible. We used best judgment to identify the publicly accessible data sets. Most articles use data from paper publications (journals, magazines, newspapers, books, reports, etc.) and we considered all of those as public. Some referenced unpublished work or data sets, some referenced data available on request, and some referenced private data. We considered all of those sources as nonpublic. For macroeconomic papers, we used common sense on what data would have been publicly available in 1960, but that might be an overestimate.

Video A1. Implementing Increased Transparency and Reproducibility in Economics

Visit the web version of this article to view interactive content.

Data Repository/Code

- The data underlying Figure 1 stem from Chetty (2012), and are available at https://doi.org/10.5281/zenodo.1453345, together with code to create the figure as presented in this document.
- Data on response rates to requests for replication materials can be found in Vilhuber (2020b), https://doi.org/10.5281/ZENODO.4267155
- Data for Table 3 are derived from Kingi et al. (2018).
- Figure 2 is sourced from Vilhuber et al. (2020), data are available at https://doi.org/10.3886/E117884V1, together with code to create the figure as presented in this document.

Footnotes

1. Unit testing is a concept from software engineering, where components of a larger piece of software are tested to ascertain that the software performs as intended. ↩

2. In fact, some define these terms in exactly the opposite way (McCullough et al., 2006). ↩

3. For an interesting overview of the history of the U.S. Census Bureau, with reference to various other related agencies that produce public statistics, see M. J. Anderson (2015). ↩

4. Similar later efforts to consolidate and standardize dispersed electronic records, such as the U.S. censuses, the Current Population Survey, and others, led to the creation of the Integrated Public Use Microdata Series (IPUMS) in 1995 (Sobek & Ruggles, 1999). ↩

5. This author benefited greatly from those early cataloging efforts at Université de Montréal, where he commenced graduate studies in 1992, and often spent many hours in the physical collection of working papers collected by Féthy Mili, the Economics Department librarian and creator of BibEc. ↵

6. "I have seen a small manufactory of this kind where ten men only were employed […] make among them about twelve pounds of pins in a day" (Smith, 1776). ↵

7. See the Appendix for methodology and data for the *AER*, *JPE*, and *ReStat*. ↵

8. The author worked with algorithms and code shared informally among labor economists as a research assistant in the early 1990s, and informally, this seems to have been standard practice. ↵

9. Prior to 2001, the Stata Technical Bulletin (STB) fulfilled the role of distributing software components. ↵

10. https://ideas.repec.org/s/boc/bocode.html, maintained by Christopher Baum, an economist at Boston College Department of Economics. For additional information about the creation of the SSC archive, see Cox (2010). Note that bibliographic information on SSC items is disseminated via RePEc; thus, it is one of the earlier examples of citable software components. ↵

11. There is no summary table of replications in Dewald et al. (1986. I classified their verbose descriptions into 'success' or 'failure' based on the extent of the replication success. Only two studies were perfectly replicated, others had some (minor) deviations, and three failed to replicate. ↵

12. The two studies cited differ in their time frames (2013 and 2011, respectively), how economics journals are identified (denominator) and how data availability policies are counted (numerator). ↵

13. Personal correspondence in 2018 with several editors and co-editors of the AEA's journals suggests that a very small number may, of their own accord, request and successfully obtain data and code as part of the refereeing process. When doing so, they report excellent compliance. ↵

14. Refusals included "too complicated for you," "I no longer have access to the data at my prior employer," and, sadly, "the co-author with the data has passed away". The data and report are not yet available to allow for a follow-up contact. ↵

15. For a first assessment of these efforts at the AEA, see Vilhuber et al. (2020). ↵

16. For additional reproducibility studies, see Berry et al. (2017), Duvendack et al. (2017), and Galiani et al. (2017). ↵

17. The term 'proprietary' refers to ownership. Many data sets limit access not because of they are 'owned' by somebody, but because balancing concerns of privacy and access yields a non-open-access solution in

order to provide both. The reason for limiting access is thus not ownership—many data owners and custodians provide open access to their data. ↵

18. For an example of a recent discussion of reproducibility relying on restricted access French data, see Chemin & Wasmer (2009, 2017) and Godechot (2016b, 2016a). ↵

19. For IPUMS, extracts from population samples (e.g., the 5% sample of the U.S. population census) rather than full population censuses (the 100% file) can be provided to journals for the purpose of replication. ↵

20. In personal conversation, Twitter's data team was receptive to improving transparency and reproducibility of research using Tweets. A conversation with Bureau Van Dijk's sales personnel was less conclusive. ↵

21. For instance, Barseghyan and coauthors (Barseghyan et al., 2013b) specifically foresaw the interest of potential replicators to access the confidential insurance company data they used for their 2013 article. The supplemental data package, provided first on the AER's website as a ZIP file and since 2019 on ICPSR (Barseghyan et al., 2013a), contains a user agreement, allowing interested users to request access to the data in a secure environment at the Cornell (University) Institute for Social and Economic Research (CISER). This fact is not discoverable by searching on either the AER's or on CISER's website. ↵

22. Hirsch & Schumacher (2004) show that the imputation methods of the CPS bias any conclusion with regard to earnings or wage gaps when the relevant variable (e.g., union status) is not part of the earnings imputation model, and point to prior well-cited studies that ignored that point (e.g., when measuring the union wage gap). ↵

23. Some recent examples are cited in Höffler (2017a). The seminal study of Dewald et al. (1986) makes note of one example. ↵

24. The AEA RCT Registry is managed by J-PAL and supersedes the J-PAL Hypothesis Registry. ↵

25. *Work, Aging and Retirement* (*OSF - Registered Reports*, 2018) and the *Journal of Development Economics* (Berkeley Initiative for Transparency in the Social Sciences, 2018) are some examples of economics-related journals using registered reports. ↵

26. A similar risk exists when submitting manuscripts to journals: editors and referees theoretically have the ability to hinder the publication of an idea while simultaneously appropriating the idea for their own research. Luckily, such unethical behavior seems to be quite rare. ↵

27. For instance, the AEA Registry notes "key information can be hidden until the time where the trial is completed" (AEA RCT Registry, 2020). The *Journal of Development Economics*, in implementing

registered reports, committed to publishing (accepted) registrations regardless of their findings, protecting the authors from any risk of being 'scooped' (*Journal of Development Economics*, 2019). ↵

28. Disclosure: I am involved in both of those working groups, either as a member, or as an outside expert. ↵