

CRRESS

**Conference on Reproducibility and Replicability in Economics and the Social
Sciences**

Table of contents

1	Home	4
2	Events	5
3	Contact Us	6
I	Session 1 - Institutional support: Should journals verify reproducibility?	7
4	Replication Packages for Journals: For and Against	9
4.1	Background	9
4.2	Main Thoughts	10
4.2.1	Why Should Journal Require Replication Packages?	10
4.2.2	Why Shouldn't Journals Require Replication Packages?	12
4.3	Conclusion	14
4.4	References	15
5	Comments on Reproducibility in Finance and Economics	16
5.1	Introduction	16
5.2	Code, Data, and Arms-Length Reproduction	16
5.3	Proprietary Data	17
5.4	Conclusion	18
II	Session 2 - Replication and IRB	19
6	Sex, Lies, and Data: New Models of Informed Consent	21
6.1	Introduction	21
6.2	Nonconsensual Porn and Nonconsensual Data	22
6.3	Intersectional Feminism, Queer Theory, Sex Education, and Critical Theory . .	22
6.4	New Models of Consent	24
6.5	Final Thoughts	25
6.6	References	26

III Session 3 - Should teaching reproducibility be a part of undergraduate education or curriculum?	28
7 Data Citations and Reproducibility	30
7.1 Introduction	30
7.2 Expected Proficiencies	31
7.3 Evidence of Broad Data Literacy Skills	31
7.4 Evidence of Narrow Reproducibility Skills	32
7.5 Proposed Instructional Intervention	33
7.6 Conclusion	33
7.7 References	34
IV Session 4: Reproducibility and confidential or proprietary data: can it be done?	36
8 Reproducibility with confidential data: The experience of BPLIM	38
8.1 Background	38
8.2 Main Thoughts	39
8.3 Conclusion	42
V Session 5: Disciplinary support: why is reproducibility not uniformly required across disciplines?	43
9 Reproducibility in Economics: Status and Update	45
9.1 Background	45
9.2 Main Thoughts	45
9.2.1 I. The (possibly unique) role of the American Economic Association . .	46
9.2.2 II. AEA Actions	46
9.3 Conclusion	47
VI Session 7: Why can or should research institutions publish replication packages?	49
10 Open Data and Code at the Urban Institute	51
10.1 Background	51
10.2 Main Thoughts	52
10.3 Conclusion	53

1 Home

The Conference on Reproducibility and Replicability in Economics and the Social Sciences is a series of virtual and in-person panels on the topics of reproducibility, replicability, and transparency in the social sciences. The purpose of scientific publishing is the dissemination of robust research findings, exposing them to the scrutiny of peers and other interested parties. Scientific articles should accurately and completely provide information on the origin and provenance of data and on the analytical and computational methods used. Yet in recent years, doubts about the adequacy of the information provided in scientific articles and their addenda have been voiced. The conferences will address the following topics: the initiation of research, the conduct of research, the preparation of research for publication, and the scrutiny after publication. Undergraduates, graduate students, and career researchers will be able to learn about best practices for transparent, reproducible, and scientifically sound research in the social sciences.

2 Events

Upcoming and past webinars can be found [here](#)

3 Contact Us

For more information, or if you are a presenter and have questions, please [contact us](#).

CRRESS is managed by co-PIs **Lars Vilhuber and Aleksandr Michuda** (Cornell University).

The organizing committee is composed of **Vilhuber, Michuda, Ian Schmutte (UGA), and Marie Connolly (UQAM)**.

Support is provided by Sara Brooks (Cornell University) as well as the staff at the Cornell University ILR School.

Part I

Session 1 - Institutional support: Should journals verify reproducibility?

Different journals have different approaches towards enforcement of their data availability policies, ranging from a thorough and complete verification including running code and checking the output, to a cursory review of the files provided to make sure they appear satisfactory, to simply receiving the data and code package and archiving it on a website or a repository. What drives the choice of approach? What are the reasons behind such choices?

In this webinar, held on September 27th, 2022, and moderated by Lars Vilhuber, we had three panelists who are experts on this topic:

1. Guido Imbens, Professor of Economics at the School of Humanities and Sciences; Senior Fellow at the Stanford Institute for Economic Policy Research; Coulter Family Faculty Fellow at Stanford University, and editor of *Econometrica*,
2. Tim Salmon, Professor of Economics at Southern Methodist University and the editor of *Economic Inquiry*, and
3. Toni Whited, Dale L. Dykema Professor of Business Administration at the Ross School of Business at the University of Michigan and editor-in-chief at the *Journal of Financial Economics*.

4 Replication Packages for Journals: For and Against

It is vital to the integrity of our field to push for greater transparency in the research that we produce. In empirical work there is a substantial opportunity to hide, even unintentionally, very subtle but important details of a project in a long series of decisions regarding how to clean and merge data sets, how to calculate variables, how to calculate summary measures and how tests are actually performed. If all it is possible to see of a paper is a set of final tables, it is often quite difficult to understand exactly how authors achieved those results and to verify that they were produced accurately. By having authors make all of those decisions transparent and available for review it allows for the possibility of others to validate the work that has been performed and it allows for future researchers to be better able to build off of that work to advance our understanding of the issues. Journals can and should facilitate that process by requiring authors to provide full details about all empirical work performed. This is not an uncontroversial viewpoint and so in what follows I will talk about some of the key arguments for and against that position.

4.1 Background

I have been on the editorial boards of many different journals for over 10 years. That experience, and my experience trying to publish in journals for much longer, has made me frequently question the editorial process, how to improve it and how journals can maintain high standards for work which they publish. In July of 2021, I took over as Editor of *Economic Inquiry* and was then in position to begin putting in place some policies which I thought would be beneficial in this regard. One of the first policies that I began working on was a policy requiring authors to share data and code related to papers published in the journal. I, of course, borrowed liberally from other journals which had already adopted such policies as there were many good models out there to borrow from. When the policy was finalized, we had chosen to fund a repository on OPENICPSR for both journals operated by the Western Economic Association International (*Contemporary Economic Policy* being the other journal) and establish a policy that requires all papers published by EI which include data analysis to publish a replication archive on that or a suitable alternative site. I had many discussions along the way to arrive

at that policy and here I will explain some of the considerations which helped me to make the final choice.

4.2 Main Thoughts

4.2.1 Why Should Journal Require Replication Packages?

We can first examine the case in favor of journals operating data archive sites like ours or in general of requiring authors to post replication packages which will allow others to reproduce their work. The main point behind this push for reproducible science is that such measures are necessary not just to maintain the credibility of individual research papers but to maintain the credibility of all academic research. There have been many examples of fraudulent work being published in academic journals over the years including many cases of researchers faking data. Two of the more famous incidences of this type of fraud were by Michael LaCour and Diedrik Stapel. In the case of LaCour, he was able to publish a paper in *Science*, supposedly the top journal across all disciplines for academic research, in 2016 which claimed to show that contact with a homosexual individual improved one's support for gay marriage proposals.¹ This was a blockbuster finding picked up by many news outlets. It was quite humiliating to many involved when it was later discovered that the data were faked. Diedrik Stapel is a repeat offender on this issue as he was able to publish many different studies in high quality journals on the basis of faked data.² There are also other types of poor quality research that are fraudulent despite using real data which show up in journals as well. Among the more notorious offenders here would be Brian Wansink, the former head of a large research center at Cornell, who was also forced to retract many articles once the methods behind those articles were revealed.³ In his case, the data existed but he engaged in methods to achieve his results which involved, to quote Cornell's Provost at the time Michael Kotlikoff, "misreporting of research data, problematic statistical techniques, failure to properly document and preserve research results, and inappropriate authorship."⁴ Many of the results from these papers had also been picked up in the popular press and so the findings of research misconduct here were quite public and embarrassing to all of the research community that allowed this work to publish. Many more examples of these problems can be found on <https://retractionwatch.com/> and indeed the fact that such a website exists is a testament to the fact that far too much problematic research somehow makes its way to the pages of scientific journals.

We clearly need to do better and requiring more transparency in empirical work at journals is a good start. Facing requirements to provide the all of the underlying data, explicit details on

¹<http://retractionwatch.com/2015/05/20/author-retracts-study-of-changing-minds-on-same-sex-marriage-after-colleague-admits-data-were-faked/>

²<https://www.apa.org/science/about/psa/2011/12/diederik-stapel>

³<https://retractionwatch.com/2022/05/31/cornell-food-marketing-researcher-who-retired-after-misconduct-finding-is-publishing-again/>

⁴<https://statements.cornell.edu/2018/20180920-statement-provost-michael-kotlikoff.cfm>

methods for data collection and code for conducting the regressions would undoubtedly deter most of the cases discussed above and many other besides. This is because being required to produce the data and make it visible to others would often unmask the underlying fraud quickly and easily. There would also be a clear public record one could check to determine legitimacy of the work. Knowing it will be harder to pass through, one hopes fewer would try and when those few still try, it should be easier to uncover the problems and deal with them as necessary. Further, not only should these requirements reduce these egregious cases of fraud, which thankfully are not that wide spread, but they will force all authors to think very carefully through their empirical processes knowing that they will be publicly viewable. This increased scrutiny should hopefully improve the quality of all research published in our journals. Preventing cases of fraud while making the details of high quality research transparently available should be a substantial boost to the legitimacy of all of our work.

It is also important that journals have policies about data availability because the ability for future researchers to reproduce existing work is necessary for the advancement of science. In many cases, one research group may wish to build upon the work already published in a journal. A first step in that process is often reproducing the initial work so that the researchers can start from there and build up. Unfortunately, if these data availability policies are not in place it is often quite difficult for a set of researchers to back out exactly what others did from a published paper alone. In one case at my own journal, a paper was submitted which was attempting to do exactly this of building off of a previous paper published at the journal. The new paper's goal was to improve on the estimation process of the previous one. The problem is that the new researchers could not reproduce the original results and so their "replication" estimation generated a result not just quantitatively different from the original authors but qualitatively different in a very meaningful way as well. This makes it then difficult to evaluate whether their improvement to the original estimation approach yielded an improvement as it is unclear that they replicated the original one correctly. That is a problem for the researchers who previously published their work as it is harder for others to build on it and it is certainly frustrating for the later researchers who cannot replicate the prior work. Having replication packages accompanying published papers can resolve this problem quickly as researchers who wish to build off of the work of others can see exactly what they did to get those results without guessing and potentially failing to identify exactly what they did.

A great example of the reason that replicating the work of others is often difficult is contained in Huntington-Klein et al (2021). This study examines the problem of replication at a deeper level than what journals usually engage in. The authors of this paper asked several teams of researchers to take the same raw data as two published papers and try to provide an answer to the same research question posed in those papers. This meant that the new researchers had to take the initial data, make all of the choices empirical researchers have to make about processing that data and specify a final regression to examine the issue. The results were that the original results often did not replicate. In some cases, the replication studies found a different sign on the key effect in question while in others, the magnitude and standard error of the effect were quite different. Importantly, in all cases, the final number of data points considered differed despite all studies starting with the same raw data with the same number of

observations. The discrepancy in the final results may have been due to the fact that different research teams often made very different choices along the way to the final specification. And thus, to really know how a team of researchers arrived at a set of results, one really needs to know more than just what was the nature of the regression conducted but you need to know all the small steps along the way to get there. Without this detailed level of information, it can be impossible to really understand how two different studies arrived at different outcomes.

It is important at this point to distinguish between two very different, though related, goals of the data availability policies of journals and how data archives may be vetted by journals. The most commonly discussed check that journals may wish to perform about a replication archive is whether one can use the archive to reproduce the results in the paper. Such a certification verifies that indeed when code is run that the results of that code reproduce what is in the paper. This verification is valuable, but a certification that the authors can re-produce their own results is not really all that useful on its own. What the paper just discussed points out is that we also need the replication sets to provide all of the details regarding how the empirical analysis was performed so that future researchers can know exactly what the authors did. With this knowledge, future researchers can begin from more robust baselines regarding published work. Without this information, we run the risk of having many parallel research programs generating seemingly conflicting results with no way to clearly determine if the conflict is due to regression specifications, different choices in data processing, errors in data processing or something else along the research chain. When designing data availability policies, we need to keep both of these goals in mind and when certifying archives as being of high quality, we need to ensure that both of these goals can be achieved.

4.2.2 Why Shouldn't Journals Require Replication Packages?

While I find the arguments above convincing regarding why journals should require replication packages, when I was contemplating putting one in place for EI, I did talk to many people who were of the opinion that journals should not be putting these requirements in place. It is worth examining their arguments against these policies to determine how convincing they are.

The first concern many would suggest about these archives is that if authors are required to post their data and their code for conducting their analysis, then others would be able to copy their work. Their concern is that the authors may have spent a great deal of time figuring out how to find the data involved, merge multiple data sets and clean them so that they work together. It may have also taken a great deal of time to implement the empirical methodology for the model in the paper. Many researchers may wish to keep that work for themselves so that they may continue to exploit that in future publications and do not want to allow others to make use of their efforts. At face value, this argument may seem somewhat convincing. While I had my own response to this, I have to say that the most convincing counter-argument against this line of thinking came from Guido Imbens in our panel discussion on this topic. He pointed out that allowing empirical researchers to hide their methods like this is similar to

allowing theorists to publish theorems while keeping the proofs hidden. A theorist could mount the same argument that the proof may have taken a long time to work out, perhaps requiring the development of special techniques in the process and they may wish to be the only ones exploiting their methods in future work. We do not allow theorists to avoid providing proofs because we need to see verification that the theorems are valid. We do not simply trust them blindly. Yet empiricists who wish to hide their methods are asking journals to blindly trust them. That should not be how publishing works. Also, while yes, making your methods and data transparent may allow others to "copy" your work, the proper way to see that is that it allows others to build off of your work. Your work can now form the foundation of the work of others and have greater impact. I would argue that the possibility that it allows others to learn more from your work is in fact one of the main reasons why journals should be requiring these packages. It is not a downside.

Another common concern about journals requiring replication packages is the suggestion that these requirements place an undue burden on authors. This can be of particular concern to certain journals as putting such requirements in place could potentially decrease the number of submissions to the journal as authors decide to submit to peer journals without such requirements. Journals likely do need to weigh this concern when considering how stringent to make their data availability policy. It is worth noting that as more and more journals adopt these policies, authors will have fewer places to submit where they can avoid these requirements and so over time concerns over this issue should diminish. It is also worth considering as a journal editor whether you want to be among the last journals not enforcing these requirements. If you are, this will mean that all those people who do not want to engage in transparent research practices will submit to your journal. As an editor, do you want to be the recipient of those submissions? Perhaps not though that decision may depend on the peer journal group for a specific journal. For journals whose peers are not yet putting these policies in place, then even high quality authors might wish to avoid the burden if they have good alternatives. For journals whose peers mostly have these requirements, then being one of the few that do not poses significant risk to the journal of receiving work for which there is a reason the authors wish to avoid transparency. Different journal editors may examine this issue and come to different conclusions on the right policy for their journal at a specific point in time. For EI, we have had the policy in place for a little less than one year and based on the current data our total submissions are slightly lower than the previous few years at this point in the cycle.⁵ There are a few other possible explanations so it is not clearly attributable to this policy but the decrease is not at a problematic level even if the data policy is responsible for the entire decline.

My other view on the issue of a replication package being a burden on authors is that this is only true if authors wait until the end of the publication process to think about the reproducibility of their work. If authors have engaged in their work in a haphazard way prior to acceptance, then it can indeed be a substantial burden to go back and document all of the data manipulation

⁵I note that in the discussion I think I said our submissions were more or less unchanged. I re-checked the data after and with the most up to date numbers we have had a small but noticeable drop.

that was done and script all of the regressions performed. If, however, authors begin thinking about these issues when they begin their research, there is no real burden and in fact I would argue that engaging in your research in a way from the beginning which will make the work replicable will actually save the authors time and allow them to do higher quality work. In my own work, I admit that early in my career I did much of my data work by hand. Then when I got a referee comment suggesting a different way to conduct a regression I would have to engage in some forensic econometrics to first back out what I actually did to get the prior result. This was wasted time and not the best way to do research. Now that I have all the analysis scripted, making changes like this is much faster and I do not have to wonder exactly how I created a variable or exactly which observations may have been dropped or why. All of that is in the scripting files from the beginning. As authors begin to expect to face these requirements and learn how to take this into account from the beginning of their analysis, the burden of providing a replication package upon acceptance of the paper diminishes substantially. I expect that these practices should be becoming more common in the profession and so the concern over this element should diminish with time. We can further diminish them by making sure that replicable research is brought into Ph.D. training programs.

A final notion that some suggested to me is that there is no need for journals to require replication packages. Individuals who want to provide their data can do so on their own sites and if there are professional incentives to do so perhaps in the form of these packages being seen as signals of high quality, everyone will do this anyway. Perhaps this could be true but most do not currently publicly archive replication files absent journal requirements. Were that to start, then it could be seen as a high-quality signal when someone does it which would mean that as journal editors we should be taking it into account in our decisions whether someone provides the data archives. If we do that, it is just a backdoor way to require replication archives but with a serious downside. If we make an accept decision based on an author saying that they will post an archive, after the paper is published authors could quickly pull that archive. Essentially, this approach is not an effective way of accomplishing the goals of research transparency. In order to ensure that the data remains available, it is best that journals maintain the archives for integrity of the process so that authors cannot manipulate the archive after the paper is published.

4.3 Conclusion

I believe quite strongly in the need for transparency in research. In order to preserve and maintain the integrity of all of academic research, we need to push for ever greater transparency in how research is done. That way when there are questions about the legitimacy of a claim, those questions can be quickly and easily addressed. This level of legitimacy is a benefit to us all. The main "cost" (if one sees it that way) would be that greater transparency limits the ability of people to publish ill-founded results. It is true that greater transparency does place greater requirements on researchers to engage in more careful and rigorous work which can

survive the greater scrutiny possible with the increase to transparency. I see this as a clear benefit rather than as a cost.

Of course, the path to this greater transparency norm will not be direct and not all journals will adopt the same standards at the same time. There are some journals leading in this direction, some following and some lagging behind. There are good reasons for different journals to be in each of those stages. As journals collectively move along this path it is important that we do so in ways that are not unduly burdensome on authors. This means that while requiring replicable archives is a valuable thing, it does not make sense for different journals to impose very different and idiosyncratic requirements about file structures and things like that such that when authors prepare a replication archive they must do a great deal of work to change it from a format suitable to one journal versus another. As a journal editor I appreciate the work done by others to establish clear guidelines on these issues which other journals can adopt as well to try to harmonize these requirements where we can.

4.4 References

Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J.R., Burli, P. et al. (2021) The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59: 944– 960. <https://doi.org/10.1111/ecin.12992>

5 Comments on Reproducibility in Finance and Economics

5.1 Introduction

Reproducibility is defined as obtaining consistent results using the same data and code as the original study. Most of the discussion of reproducibility has centered around the many obvious benefits. Reproducible research advances knowledge for several reasons. It reduces the risk of errors. It also makes the processes that generate results more transparent. This second advantage has an important educational component, as it helps disseminate not just results but processes. However, reproducibility is not without costs. Good research procedures consume resources both in terms of a researcher's own efforts and in terms of the involvement of arms-length parties in actually reproducing the research. This second cost is not just a time cost; it is pecuniary as well.

Thus, reproducibility is a good that is costly to produce and that has many positive externalities. Researchers internalize many of the benefits of reproducibility, especially in terms of research extendability and personal reputation. However, they do not internalize any of the benefits to the research community at large. Because reproducibility is costly, it is unlikely to be produced at a socially optimal rate by any individual researchers. Thus, the questions are the extent to which reproducibility should be subsidized and who should subsidize it. Should all research be reproduced by arms-length parties, and what are the least costly policies that facilitate reproducible research? The rest of this note is organized around policies regarding actual reproduction and proprietary data.

5.2 Code, Data, and Arms-Length Reproduction

One low-cost and easily implementable set of policies that enhances the reproducibility of research is journals' data and code disclosure policies. In the age of inexpensive data storage and an abundance of public repositories, the costs of these policies are small, and the policies should be implemented. They impose some costs on researchers in terms of organizing data and code, but well-organized data and code are already an essential part of the research process, so these costs should be small.

While simple to implement, this low-cost policy is not without non-pecuniary drawbacks for journals. The code and data can be incomplete, poorly documented, or unusable. Moreover, journal editors have to retract articles that, after publication, cannot be reproduced. In economics, these concerns have prompted journals to start arms-length reproduction of results. The benefit of this policy is primarily that authors and journals can be confident that the code submitted with an article actually works to reproduce the results.

However, the pecuniary costs of this policy can be substantial. It is expensive for journals to hire data editors and well-trained research assistants, and many academic journals run on tight budgets. It is often time-consuming for authors to comply with reproducibility requirements. This last issue is particularly burdensome for authors who cannot afford research assistance.

While the above issues involve costs, the following are more fundamental. Reproducibility policies give researchers incentives to do research that is easier to reproduce, thus restraining research innovation that requires either large data or intense computing. Most importantly, code that can run on data and reproduce results can still contain errors.

These arguments imply that while individual researchers are likely to underproduce reproducibility, it is also unlikely optimal for the progress of science that all research be reproduced before publication. Some papers, even those in the very best journals, rarely get read or cited, and the benefits of reproducing these papers are small.

However, ex-ante, it is hard to know which papers will attract attention and which will not. One solution that lies between data and code disclosure and arms-length reproduction is verification. It is much less expensive to verify the contents of a replication package than to do an actual reproduction. Verification might consist of checking for the existence of replication instructions, an execution script, or either data or pseudo-data. This type of service could be provided by journals or other third parties, much as copy editors fix syntax and grammar errors before articles are submitted. At that point, reproducibility would be left up to the academic community, with the more important pieces of research being subject to greater scrutiny.

A final issue with reproducibility is education. In economics and finance, students are not taught how to create reproducible research. An improvement that would go a long way toward improving the culture surrounding reproducibility would be to teach PhD students how to organize research projects and to write code in such a way that others can reproduce results easily. This type of education would lower the costs to individual researchers of making their own research reproducible.

5.3 Proprietary Data

A possibly larger challenge for reproducibility than verification or arms-length execution of code is proprietary data. A clarification is necessary because not all types of data with restricted access are completely secret, that is, available only to the data provider and a researcher. For

example, commercial data sets are not secret, just costly to obtain. Similarly, administrative datasets are not secret. They just require special permission. In contrast, proprietary data cannot be offered to the research community at large for the purposes of reproducing the results. So the question is whether journals should discourage the use of this type of data or require that verifiers have access to the data. Given the large number of studies using proprietary data, this issue is possibly more important than the issue of running code.

5.4 Conclusion

In conclusion, the reproducibility of research is essential for the advancement of science. However, it is not without costs, so blanket statements that all research should be reproducible are not feasible. Instead, feasible policies include those that lower the costs for others to replicate research. Data and code disclosure is a low-cost policy that should be implemented widely. Verification of code and data packages is a slightly more costly option. Arms-length reproduction is a much more costly alternative. Finally, perhaps the most important issue that impedes reproducibility in finance and economics is the use of proprietary data.

Part II

Session 2 - Replication and IRB

One of the most crucial dimensions that Institutional Review Boards are interested in are the protocols that researchers have in place to protect their subjects' privacy. This often leads to researchers writing in their IRB protocols that they will destroy their data once their project is complete. Understandably, however, destruction of data makes it impossible to verify and replicate work, which is increasingly becoming a vital part of modern science. How should data privacy be handled in the wake of the replication crisis? What protocols and standards should be put in place to minimize the risk of data leakage? Or should data be destroyed after some time span?

6 Sex, Lies, and Data: New Models of Informed Consent

This paper identifies gaps within the current research ethics regime that can be complemented by intersectional feminism, queer theory, sexual education, BDSM, and critical theory. When these disciplines are applied, informed consent models can better integrate revocability, ongoing consent, and contextual consent. Future work on informed consent and research ethics should incorporate these areas to preserve the dignity of research subjects and equitably benefit all members of the research process.

6.1 Introduction

In the summer of 2015, a company called Ashley Madison was hacked and over seven million users' data was leaked (Lord, 2017). Data breaches like this are unfortunately common, but as Ashley Madison was specifically created to facilitate infidelity, the social fallout of the breach was more harmful than most. After the breach and subsequent fall-out, which included many individuals being publicly named and shamed after their information was found in the customer records, many users of the site ended up getting divorced, some lost their jobs, and there were multiple confirmed reports of self-harm (Baraniuk, 2015).

Several researchers saw the data breach as an opportunity to advance their research. They used the leaked data to explore different questions about user demographics, geography, and risk-taking behavior and published the results in multiple peer-reviewed articles (Billau, 2017; Vedantam, 2016). However, in order to be published, these studies first had to be approved by their university's Institutional Review Board (IRB).

IRBs are intended to reduce the potential risks and harms in research proposals that involve human subjects (or data thereof) and are often perceived as a marker of credibility for ethical research. All IRBs use something called the Common Rule, which is a set of federal guidelines that dictate how to conduct ethical research, and there is a stipulation in it that says if data is already public, then there is little potential for risk or harm for it to be used in research. It makes no stipulations about how that data became public. In fact, IRBs have no remit to review secondary uses of data, with bio-specimen data as the sole exception. To an IRB, public

data is considered low risk regardless of how that data was obtained, and regardless of how sensitive the data is, and evaluating its use is not within their scope of review. Researchers got their studies approved by their IRBs, and academics continue to use and cite Ashley Madison data in publications. As a result, many more people see the leaked data and names of the implicated humans than would have had those papers never been published.

6.2 Nonconsensual Porn and Nonconsensual Data

There are parallels between the use of leaked data sets like the Ashley Madison hack, and the viewing of nonconsensual pornography (NCP). For the unfamiliar, NCP is "the distribution of sexually graphic images or videos of an individual without their consent in the context of an intimate relationship," (Carter, 2021). Perpetrators of NCP sometimes upload photos or videos of their victims onto pornography websites, which allow many more people to view them. Like most platforms, these websites often use engagement metrics, like view counts, which help determine if a video is "popular", with more popular media receiving more promotion. From a victim's perspective, the more people who see their image or video, the worse the mental and/or emotional harm. While the original perpetrator inflicted a significant harm, that harm gets compounded every time it gets viewed or shared by other people, and by viewing it, it raises the likelihood that someone else will view it.

When researchers use data that was collected nonconsensually, unethically, or illegally, it has the potential to amplify the harm of the original data collector by pointing other people to it. Using unethical data for research is not the moral equivalent of the original hack, but there are many circumstances where it may still be considered morally wrong. The fact that the Common Rule and IRBs don't have the mechanisms necessary to address this issue means that we need more tools to fully judge the needs and bounds of ethical research (Jordan, 2022). Fortunately, there are many disciplines and scholars that have developed theoretical tools and practices we can use to improve the ethical conduct of research. I offer several areas that have been personally productive, though there are many others that can be used.

6.3 Intersectional Feminism, Queer Theory, Sex Education, and Critical Theory

The first area that has been helpful for developing a fuller understanding of consent is intersectional feminism. Feminism is an interdisciplinary approach to addressing oppression related to gender identity and expression. Intersectionality is a theory that when someone has more than one marginalized identity (e.g., being a Black woman), the combination of those identities produces a greater risk than the sum of their parts (Crenshaw, 2006). Intersectional feminism interrogates oppression on multiple fronts including gender, race, class, disability, and others, especially how these oppressions interact with each other. These theories have much to say

about what informed consent should look like, especially when it comes to bodily autonomy and agency. An intersectional feminist approach to informed consent will ensure that participants have control over their participation in a research study, individual control over their personal data, and communal control over communal data (Sterling, 2011; Fiesler, McCann, Frye & Brubaker, 2018). Informed consent can sometimes be framed as a liability waiver for institutions. Intersectional feminism would reframe informed consent as an expression of care for the wellbeing of the person which may supersede the research goals of the researchers or the legal liability of the university.

A second helpful area is queer theory. Queer theorists critique dominant social expectations of sexual orientation and gender identity (Cohen, 1997). Informed consent can sometimes mirror hetero-patriarchal models of power by framing consent as something only certain people are qualified to give (cis, heterosexual men), while others (LGBTQ people) are disqualified, for example, the FDA's ban on gay and bisexual men donating blood (Human Rights Campaign, 2020). Integrating queer theory and listening to the LGBTQ community can aid in analyzing how certain processes can unintentionally reify discriminatory ways of thinking, and they can help us move toward more inclusive ways of achieving informed consent (Edenfield, 2019; de Heer, Brown & Cheney, 2021).

A third area is sexual education and the BDSM community. Inclusive and evidence-based sex educators have developed models of affirmative and enthusiastic consent that are sensitive to context (Center for Sex Education, 2016). Bondage/Discipline, Dominance/Submission, Sadism, and Masochism (BDSM) are consensual sexual acts that involve a power dynamic between partners. This kind of sex has a higher potential for risk because there are elements of pain or power involved and the BDSM community has developed many practices around consent and communication to mitigate those risks such as safe words (Dunkely & Brotto, 2020). These practices include experiences ranging from in-person to remote relationships, sometimes involving people's bodies while other times involving domination over personal documents or computers (Vogt & Goldman, 2018). Sex education and BDSM have richer, more nuanced theories and consent practices than most researchers at universities.

A fourth area is critical theory, in particular its analysis of power and push for social change. Critical theory takes many forms and is often integrated into existing disciplines as a way to challenge structural assumptions and redistribute resources and opportunities within them, such as critical legal studies, critical pedagogy, and others. A critical approach to consent would be attuned to who has power within an informed consent transaction, who doesn't, and attempt to redress the asymmetrical dynamic. Applying a critical theory approach to informed consent can produce some unanticipated results. For example, the Panama Papers were a set of leaked documents exposing offshore financial transactions including criminal tax evasion, money laundering, and other financial crimes around the world (Fitzgibbon & Hudson, 2021). This was likely a situation where the material was obtained unlawfully and without the consent of the people whose information was publicized. Hundreds of papers and many books have been written that detail the Panama Papers and the individuals named in them. A traditional approach to informed consent would say that the Panama Papers leak was a

wrongful violation of privacy. A critical consent approach would disregard claims of privacy by the victims because of the nature of the information leaked, namely a highly organized global financial system used by the rich and powerful to hide assets at the expense of the poor and powerless. Informed consent with a critical theory lens inverts traditional power dynamics, meaning that some rules about research ethics should be broken when following them would cause inequitable outcomes.

6.4 New Models of Consent

When you look across the areas of intersectional feminism, queer theory, the LGBTQ community, sex educators, the BDSM community, and critical theory, there are commonalities that can be pulled out to help develop a better framework of consent in research. Using these as reference points, informed consent must be at least three things: revocable, ongoing, and contextual.

Revocability means that after you give your consent to something, you can change your mind at any time and for any reason. Current research practice does this somewhat. Most consent forms state that research participants can stop their participation at any time, but there are many informal and implicit forces that discourage this. Participants may feel that by changing their minds they would inconvenience or disappoint the researchers, who are in a position of authority. Some participants may be persuaded by the sunk cost fallacy that if they've done something long enough, they might as well finish it even if they would prefer not to. In general, research participant consent is revocable only during the data collection phase. Once that data is used to publish something, there is almost nothing that a participant can do to remove their data from the study.

Ongoing consent means that the 'one and done' model of most research is inadequate. In practice, it means frequent points of intentional communication gauging the interest of the participants in continuing in the research study. This does not have to devolve into performative check-ins that likely produce consent fatigue (Ranisch, 2021). The modes of ongoing consent should adapt to the environment and be naturally integrated into the requirements of the participant experience, so that providing or revoking consent is both easy and perceived as casual. In practice, this can look like introducing several points in a study where participants have to opt-in to continuing, with the expectation that if they don't, they automatically discontinue participation. Sometimes called contextual consent or just-in-time consent, this makes leaving a study mid-way through seem less interruptive or socially uncomfortable (IF, 2022).

Contextual consent means that there isn't a template or formula for getting consent that can be applied across all research studies. When people consent to having sex, it is for specific people, places, and times. Giving consent to have sex once does not mean that you consent to sex with that person at any time and place in the future. Contextual consent adapts to the specific conditions it is being offered in and adjusts how it is communicated. In practice, this could

look like developing unique consent methods for different populations that are attentive to the cultural and rhetorical differences, even if the research protocol is otherwise the same. For some populations, the standard written forms of consent using academic language may be adequate. For other populations, offering techniques using visual, auditory, narrative, behavioral, or game-based methods might be better suited to get meaningful consent

Integrating these informed consent practices into human subject research will significantly improve the experience of participants as well as reduce the potential for unethical data collection and sharing. Academia has many justifications to ignore informed consent. Most times consent is neglected, the subsequent risks are distributed unequally among participants. The people who are put at the most risk are usually the people who are already the most vulnerable: women, children, people who are LGBTQ, non-binary, disabled, poor, people of color, and many others. Researchers have a responsibility to conduct research ethically and with respect for individual privacy needs and expectations. Using these consent practices by integrating intersectional feminism, queer theory, the LGBTQ community, sex education, BDSM, and critical theory can increase the quality of research and respect the dignity of the participants.

An admittedly difficult area for applying these frameworks is when the circumstances of data collection and informed consent are more quotidian. The examples used above are dramatic when compared with most of the common research practices such as collecting public social media data from places like Facebook, Reddit, or Twitter. When the consent obtained is ambiguous, when the data being collected isn't particularly sensitive, or when the potential for significant social benefit is high, applying these theoretical tools isn't straightforward. Reasonable people can, and often do, disagree as to what ethical values are at stake and how to adjudicate them. The unfortunate ethical gray area this produces is still better than having clear but inequitable guidelines.

6.5 Final Thoughts

It is highly unlikely that any rules developed to address the growing complexities of research ethics, especially in technology spaces, will be able to address every possible ethical dilemma that arises. The four areas offered above are tools that can help with some of the gaps currently found in research ethics, but they need to be constantly reexamined and supplemented as new issues present themselves. Ethical research is a horizon, not a place; we never arrive. By prioritizing the wellbeing of the most vulnerable and marginalized people in our communities and continually inventing better models of informed consent, we can take more confident steps towards that horizon.

6.6 References

Baraniuk, C. (2015). *Ashley Madison: 'Suicides' over website hack*. BBC. August 24.

<https://www.bbc.com/news/technology-34044506>

Billau, C. (2017) *Academic research uses hacked Ashley Madison data to map areas with most cheating husbands*. UToday, Alumni, Arts and Letters. January 10.

https://news.utoledo.edu/index.php/01_10_2017/academic-research-uses-hacked-ashley-madison-data-to-map-areas-with-most-cheating-husbands

Carter, C. (2021). *An Update on the Legal Landscape of Revenge Porn*. National Association of Attorneys General. November 16.

<https://www.naag.org/attorney-general-journal/an-update-on-the-legal-landscape-of-revenge-porn/>

Center for Sex Education, (2016). *Why Comprehensive Sex Ed and Consent Education Go Hand in Hand*, May 13.

<https://www.sexedcenter.org/why-comprehensive-sex-ed-and-consent-education-go-hand-in-hand/>

Cohen, C. J. (1997). Punks, bulldaggers, and welfare queens: The radical potential of queer politics? *Glq*, 3(4), 437-465.

<https://doi.org/10.1215/10642684-3-4-437>

Crenshaw, K. W. (2006). *Intersectionality, identity politics and violence against women of color*. *Kvinder, Køn & Forskning*, (2-3)

<https://doi.org/10.7146/kkf.v0i2-3.28090>

de Heer, B., Brown, M., & Cheney, J. (2021). Sexual consent and communication among the sexual minoritized: The role of heteronormative sex education, trauma, and dual identities. *Feminist Criminology*, 16(5), 701-721.

<https://doi.org/10.1177/15570851211034560>

Dunkley, C. R., & Brotto, L. A. (2020). The role of consent in the context of BDSM. *Sexual Abuse*, 32(6), 657-678.

<https://doi.org/10.1177/1079063219842847>

Edenfield, A. (2019). Queering consent: Design and sexual consent messaging. *Communication Design Quarterly Review*, 7(2), 50-63.

<https://doi.org/10.1145/3358931.3358938>

Fiesler, C., McCann, J., Frye, K., & Brubaker, J. R. (2018, June). Reddit rules! characterizing an ecosystem of governance. In *Twelfth International AAAI Conference on Web and Social Media*.

<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/viewPaper/17898>

Fitzgibbon, W., Hudson, M. (2021) Five years later, Panama Papers still having a big impact. International Consortium of Investigative Journalists. April 3, 2021

<https://www.icij.org/investigations/panama-papers/five-years-later-panama-papers-still-having-a-big-impact/>

IF. (2022) “Giving and Removing Consent: Just-in-time Consent”. Data Patterns Catalog. IF

<https://catalogue.projectsbyif.com/patterns/just-in-time-consent/>

Human Rights Campaign. (2020) Blood Donations.

<https://www.hrc.org/resources/blood-donations>

Jordan, S. (2022) The Playbook: Data Sharing for Research. The Future of Privacy Forum.

<https://fpf.org/wp-content/uploads/2022/12/FPF-Playbook-singles.pdf>

Lord, N. (2017). *A Timeline of the Ashley Madison Hack*. Digital Guardian. July 27, 2017.

<https://digitalguardian.com/blog/timeline-ashley-madison-hack>

Ranisch, R. (2021). Consultation with doctor twitter: Consent fatigue, and the role of developers in digital medical ethics. *American Journal of Bioethics*, 21(7), 24-25.

<https://doi.org/10.1080/15265161.2021.1926595>

Sterling, R. L. (2011). Genetic research among the Havasupai: A cautionary tale. *AMA Journal of Ethics*, 13(2), 113-117.

<https://journalofethics.ama-assn.org/article/genetic-research-among-havasupai-cautionary-tale/2011-02>

Vedantam, S. (2016). *Ashley Madison Hack Inspires Social Scientists to Look Behind the Names*. National Public Radio. April 28.

<https://www.npr.org/2016/04/28/476060486/ashley-madison-hack-inspires-social-scientists-to-look-behind-the-names>

Vogt, P. J., Goldman, A. (2018) Episode 116 Trust The Process. /reply-all/. February 28, 2018.

<https://open.spotify.com/episode/0NGOihC8u0GBuxdV3fEBLb>

Part III

Session 3 - Should teaching reproducibility be a part of undergraduate education or curriculum?

This session was held at the Southern Economics Association meeting on November 20th, 2020. Panelists discussed teaching reproducibility (TIER Protocol), the involvement of undergraduates for replications based on restricted-access data, and other topics.

7 Data Citations and Reproducibility

in the Undergraduate Curriculum

Data citations are the foundation of reproducibility. To develop reproducibility skills among undergraduate students we must start with basic data literacy skills such as citing data consistently.

Disclaimer: The views expressed in this article are those of the authors and don't necessarily reflect the position of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

7.1 Introduction

The scholarship of teaching and learning in economics documents multiple efforts to bring the quantitative dimension of our professional work closer to the undergraduate college curriculum.

Economics educators describing data-focused assignments and projects (Wolfe, 2020; Halliday, 2019; Wuthisatian and Thanetsunthorn, 2019; Marshall and Underwood, 2019; Mendez-Carbajo, 2015 & 2019) highlight the data-finding step of these projects. Even when the datasets are directly provided to the students, (e.g., Easton, 2020) the instructors emphasize the broader literacy dimensions of the assignments. However, there is neither professional consensus about how to build data-literacy skills (Wuthisatian and Thanetsunthorn, 2019) or much actual research on their mastery among economics students (Halliday, 2019).

In this chapter, we document baseline proficiency levels among undergraduate college students related to identifying data series and their sources. We also put forward an accessible pedagogical strategy to develop basic reproducibility skills.

We argue reproducibility should be part of the undergraduate curriculum in economics because it is a valuable professional skill to be developed throughout the curriculum by consistently citing the data sources used in economic arguments. We must instill the practice leading by example and enrolling the help of librarians

7.2 Expected Proficiencies

There is a natural overlap regarding the development of data-literacy skills between economics and library science: both disciplines value it and contribute to its development.

The two seminal descriptions of data literacy expected proficiencies among undergraduate students are provided by Hansen (2012) and Pothier and Condon (2019). The first of the seven broad competencies of economics majors named by Hansen directly address data provenance. It states: “Access existing knowledge: [...] Track down economic data and data sources. Find information about the generation, construction, and meaning of economic data.”

The library science perspective provided by Pothier and Condon is articulated through seven expected data competencies of economics and business majors. The last one states: “Data ethics: The principles of data ethics are built on data ownership, intellectual property rights, appropriate attribution and citation, and confidentiality and privacy issues involving human subjects.”

The utilitarian and ethical aspects of data reproducibility outlined above are bridged by the American Economic Association’s (AEA) (2020) Data and Code Availability Policy, which clearly states “All source data used in the paper shall be cited, following the AEA Sample References.” However, the scholarship documenting the collaboration in this area between instructional economics faculty and librarians is limited. Neither the calls by economics instructors (McGrath and Tiemann, 1985; Li and Simonson, 2016; Mendez-Carbajo, 2016) nor the experiences documented by librarians (Wheatley, 2020; Wilhelm, 2021; Waggoner and Yates Habich, 2020) appear to have broad impact.

7.3 Evidence of Broad Data Literacy Skills

Mendez-Carbajo (2020) documents baseline levels of data literacy competency in several areas key to the accurate and ethical use of data for communication and decision-making among high school and college students.

In the online economic education module produced by the Federal Reserve Bank of St. Louis “FRED Interactive: Information Literacy”, two separate groups of high school students ($N=450$) and college students ($N=912$) answer seven pre-test questions. The questions are mapped to the data literacy competencies described by both Pothier and Condon (2019) and Hansen (2012).

The analysis finds effectively identical levels of average baseline data literacy competency between high school and college students. However, it also documents much higher levels of perceived self-efficacy among college students than among high school students. In other words, college students are no more knowledgeable or skilled than high school students but

are significantly more confident in their work. This finding highlights a major challenge for instructors working to develop the expected proficiencies identified in the literature: the average college student is unduly comfortable in their limited understanding of the primary sources of economic data.

7.4 Evidence of Narrow Reproducibility Skills

During the fall semester of 2020, we distributed a short online assignment to all 854 students enrolled in two different upper-division economics courses offered by a large public university in the United States.

On average, the students are slightly above 20 years of age, 49% identify themselves as female, 21% identify as non-White racial or ethnic minorities, and 92% report English is their native language. Academically, 87% of students are business, economics, or finance majors and hold a grade point average of 3.41. Also, 68% of students are currently enrolled in a statistics course required by their program and, on average, have previously completed more than one and a half economics courses.

The assignment had three sections:

- First, the students were directed to read a brief, 900-word, essay on [how to create data citations with FRED®](#). This essay provided background on the value of good data citations for practitioners of economics and could be used as reference material for the next two sections of the assignment.
- Second, the students were directed to read two short --under 600 words, economic essays. See them [here](#) and [here](#). Each included a line graph of economic data. In the text, the authors referenced the data series and their sources while interpreting the quantitative information presented in the graph.
- Third, the students were asked to complete three tasks: identify the data series discussed in the essay; identify the sources of the data series discussed in the essay; and identify the missing elements of a data citation provided in the essay.

The assignment was completed in its entirety by 501 students. Table 1 reports our findings.

Table 1. Data Literacy Skills

Scores, Misconceptions and Errors	Essay A	Essay B
Identifies Series Correctly	0.57	0.47
Identifies Sources Correctly	0.21	0.03
Identifies Incomplete Citation	0.18	-0.04
Can't Identify Sources	0.05	0.12
Confuses Source with Distributor	0.72	0.73

Scores, Misconceptions and Errors	Essay A	Essay B
Considers Citation to be Complete	0.25	0.40

Note: Data Literacy Scores:
$$= (\# \text{ correct} - \# \text{ incorrect}) / (\# \text{ total})$$

We document very weak student data literacy competencies associated with narrow reproducibility skills. Data literacy scores related to correctly identifying the sources of the data or recognizing an incomplete data citation are very low. Moreover, we document a frequent misconception of confusing the data source with the distributor.

These findings have practical implications for instructors, whether they are librarians or economic educators. Our work suggests there is a substantial instructional opportunity to help students develop the ability to recognize data series and their sources. In that regard, disambiguating the roles of data distributors and data sources can potentially yield large benefits to students, who would be able to acquire a more sophisticated understanding of how data are created and made available.

7.5 Proposed Instructional Intervention

We propose a broad instructional intervention for economics instructors reflecting the fact that correctly citing the data is a foundational literacy skill.

- Lead students by example and consistently name the sources of all data referenced or used in your teaching.
- Embed this practice in all your teaching, regardless of the type or subject of the course.
- Enroll the help of librarians by leveraging their ongoing instructional outreach on information literacy to include data citations.

Proficiency in identifying data sources is foundational to the development of reproducibility skills. The earlier and the more frequently students are exposed to best practices in data citations, the more effortlessly they will be able to adopt sophisticated professional replicability practices.

7.6 Conclusion

Reproducibility should be part of the undergraduate curriculum in economics:

- It is a valuable professional skill that shows the background work that goes into doing economic research. Citing the sources of the data makes research work more thorough.

- This skill should be developed throughout the curriculum. This skill is not particular or exclusive to econometrics or statistics courses.
- The first step is to consistently cite the data sources used in economic arguments. This includes data tables, plots, and in-text references.
- We must instill the practice by leading by example. Economics educators should enroll the help of librarians in developing this skill among students.

7.7 References

American Economic Association. (2020). Data and Code Availability Policy. <https://www.aeaweb.org/journals/code-policy>.

Easton, T. (2020). Teaching econometrics with data on coworker salaries and job satisfaction. *International Review of Economics Education*, 34, 100178. DOI 10.1016/j.iree.2020.100178.

Halliday, S. D. (2019). Data literacy in economic development. *The Journal of Economic Education*, 50 (3), 284-298, DOI: 10.1080/00220485.2019.1618762

Hansen, W. L. (2012). An expected proficiencies approach to the economics major. In *International handbook of teaching and learning economics*, ed. G. Hoyt and K. McGoldrick, 188–94. Cheltenham, UK and Northampton, MA: Edward Elgar.

Li, I., and Simonson, R. D. (2016) The value of a redesigned program and capstone course in economics. *International Review of Economics Education*, 22, 48-58, DOI: 10.1016/j.iree.2016.05.001.

Marshall, E. C., and Underwood, A. (2019). Writing in the discipline and reproducible methods: A process-oriented approach to teaching empirical undergraduate economics research. *The Journal of Economic Education*, 50 (1), 17-32. DOI: 10.1080/00220485.2018.1551100

McGrath, E. L., and Tiemann, T. K. (1985). Introducing empirical exercises into principles of economics. *The Journal of Economic Education*, 16 (2), 121-127. DOI: 10.1080/00220485.1985.10845107

Mendez-Carbajo, D. (2015). Visualizing data and the online FRED database. *The Journal of Economic Education*, 46 (4), 420-429. <https://doi.org/10.1080/00220485.2015.1071222>

Mendez-Carbajo, D. (2016). Quantitative reasoning and information literacy in economics. In *Information Literacy: Research and Collaboration across Disciplines* (pp. 305-322), Barbara D'Angelo, Sandra Jamieson, Barry Maid, and Janice R. Walker (editors). Perspectives on Writing. Fort Collins, Colorado: WAC Clearinghouse and University of Colorado Press. <https://wac.colostate.edu/books/infolit/chapter15.pdf>

Mendez-Carbajo, D. (2019). Experiential learning in macroeconomics through FREDcast. *International Review of Economic Education*, 30 (1). DOI: 10.1016/j.iree.2018.05.004.

- Mendez-Carbajo, D. (2020). Baseline competency and student self-efficacy in data literacy: Evidence from an online module. *Journal of Business & Finance Librarianship*, 25:3-4, 230-243. DOI: 10.1080/08963568.2020.1847551
- Pothier, W., and Condon, P. (2019). Towards data literacy competencies: Business students, workforce needs, and the role of the librarian. *Journal of Business and Finance Librarianship* 25:3-4, 123-146. DOI: 10.1080/08963568.2019.1680189
- Waggoner, D., and Yates Habich, B. (2020). Collaboration is the key: faculty, librarian and Career Center professional unite for marketing class success. *Journal of Business & Finance Librarianship*, 25:1-2, 82-91. DOI: 10.1080/08963568.2020.1784658
- Wilhelm, J. (2021). Joint venture: An exploratory case study of academic libraries' collaborations with career centers. *Journal of Business & Finance Librarianship*, 26:1-2, 16-31. DOI: 10.1080/08963568.2021.1893962
- Wheatley, A., Chandler, M., and McKinnon, D. (2020). Collaborating with faculty on data awareness: A case study. *Journal of Business & Finance Librarianship*, 25:3-4, 281-290. DOI: 10.1080/08963568.2020.1847553
- Wolfe, M. (2020). Integrating data analysis into an introductory macroeconomics course. *International Review of Economics Education*, 33, DOI: 10.1016/j.iree.2020100176
- Wuthisatian, R., and Thanetsunthorn, N. (2019). Teaching macroeconomics with data: Materials for enhancing students' quantitative skills. *International Review of Economics Education*, 30, 100151. DOI 10.1016/j.iree.2018.11.001.

Part IV

Session 4: Reproducibility and confidential or proprietary data: can it be done?

What happens to reproducibility when data are confidential or proprietary? Many journals can only ask that detailed access procedures be provided in a ReadMe file, but what mechanisms could be used to conduct computational reproducibility checks on such data? Should authors temporarily share their data with the journal for the purposes of reproducibility verification, even if they are not part of the public data replication package? Is it feasible to use a network of “insiders” to run code provided as part of a data replication package to assess reproducibility? Could a “certified run” be used?

8 Reproducibility with confidential data: The experience of BPLIM

Access to confidential data is often seen as a major obstacle to reproducibility. The Banco de Portugal microdata laboratory (BPLIM) has been providing access to confidential data for research purposes since 2016. Based on our experience, we argue that our data access model is an opportunity to demand that researchers integrate good reproducibility practices into their research process. Our goal is to automate the production of a reproducibility package as part of the researcher's process of analyzing confidential data. We hope that this approach can overcome the limitations of working with confidential data and help certify reproducibility with confidential data. The approach we propose can also be adopted by researchers using non-confidential datasets.

8.1 Background

The Banco de Portugal Microdata Laboratory (BPLIM) was established in 2016 with the primary goal of promoting external research on the Portuguese economy by making available data sets collected and maintained by Banco de Portugal (BdP). By making this information available to researchers from around the world, BdP aims to support the development of evidence-based policies and insights that can benefit the Portuguese economy and society. However, given that some of these data sets contain highly sensitive information BPLIM had to implement a data access solution that preserved the confidentiality of the data.

The common approach by other Research Data Centers that make confidential data available for research involves the provision of on-site access to accredited external researchers in a secure computing environment. However, in the case of BPLIM this approach was deemed undesirable for two reasons. First, because it would limit access to a handful of researchers who were able to come to the Bank's premises. Second, because there were still concerns that a breach of confidentiality might occur if individuals from outside the bank could gain access to original data sets that contained confidential information.

After an internal debate at the Bank it was decided that the solution to be adopted by BPLIM had to be based on the following principles:

- access free of charge and only for scientific purposes;
- all data should be analyzed on the servers of the Bank;
- external researchers were granted remote access to the server;
- confidential datasets placed on the server had to always be perturbed/masked;
- researchers could always ask BPLIM staff to run their scripts on the original data.

The general workflow defined for data access at BPLIM was the following. After a research project is approved and the external researchers are accredited an account is opened on the BPLIM external server. External researchers gain access to a computing environment that does not allow users to transfer files to and from the server. They have access to a restricted area where standard software such as *Stata*, *R*, *Julia* and *Python* are available. Since there is no connection to the internet, installation of specific packages has to be requested from the staff. The datasets for the project are placed in a read-only folder. For the confidential datasets what is placed in the account of the researcher are perturbed versions of the data (noise is added to the original data). The researcher implements all scripts based on the data he/she has available and produces the (non-valid) outputs required for the project. Once researchers complete this task, they can ask BPLIM staff to rerun their scripts, this time using the original confidential data. For this process to be successful BPLIM staff must first run the scripts using the same data as the researcher to verify that the scripts written by the researchers reproduce exactly the outputs (typically graphs and/or tables). This process is done in a different server (BPLIM internal server). Only upon completion of this first step can BPLIM staff modify the scripts, this time to read the original data and regenerate the intended outputs. These outputs are then subject to standard output control checks for confidential data and delivered to the researcher.

8.2 Main Thoughts

Over time we have come to realize that this somewhat cumbersome process of running the code thrice, first by the researcher on the perturbed data, second by BPLIM staff again on the perturbed data, and finally, on the original data was in fact an exercise on reproducibility. Even though the reproducibility check was on the perturbed data it was already a very good assurance that a reproducibility check would hold on the original data.

We realized that a great deal of our work involved reproducing the results obtained by the researchers with the perturbed data and that led us to look at ways to improve our workflow. It became obvious that the process could be streamlined and would be more efficient if researchers adhered to the best practices on reproducibility. Hence, as part of our strategy, we have decided to raise awareness of our researchers to the need of implementing good practices on reproducible research. We have been doing this by several means. For example, we have held practical workshops which are designed to enhance the skills of our researchers. For these workshops

we invite leading experts to present best practices and recent developments on data analysis. We also provide direct advice to the researchers, prepare templates and documentation, and make available tools that facilitate the analysis of our datasets (particularly for more complex tasks such as building a panel or calculation of specific variables).

On the other end, it was also obvious that there was margin for improvement on our work sequence. One possible improvement was the assurance that the computing environment used by the researcher on BPLIM's external server was identical to that used by BPLIM staff when reproducing the code. Thus, for the case of researchers that work with open-source software, we have been incentivizing researchers to work with Singularity containers. This facilitates our work because we are sure that our reproducibility check is implemented in the same self-contained environment that was used by the researcher. Researchers that use Stata can resort to containers but in that case, it is easier to control the environment because we install all packages on a folder that is specific to each project and have developed tools that facilitate comparison of the Stata ado files across environments. (all tools are publicly available and can be found at <https://github.com/BPLIM/Tools/tree/master/ados/General>.)

More recently, we have worked on shifting the burden of the reproducibility check to the researcher itself. We are developing an application that we are presently testing with a select number of researchers. The application is targeted mainly at researchers that use BPLIM's (perturbed) confidential datasets but we hope to eventually convince all other users to take advantage of it. To illustrate, we provide a screenshot of the application:

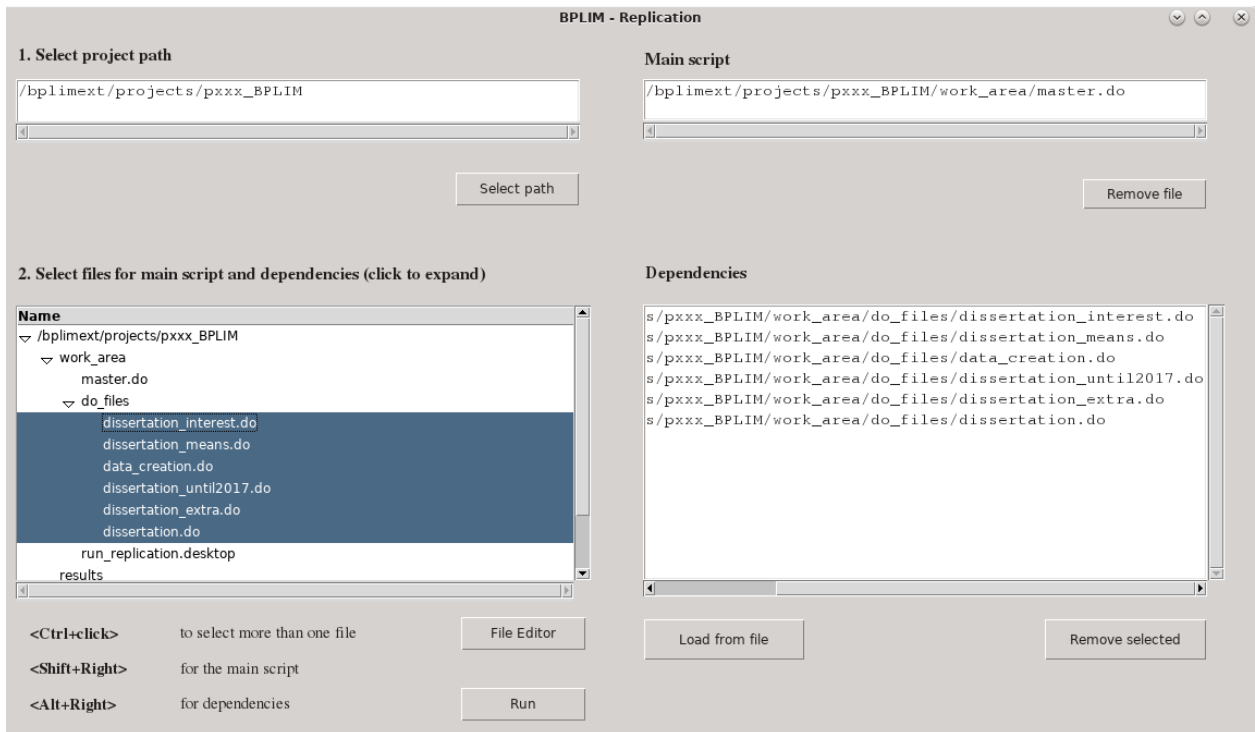


Figure - BPLIM Replication tool – selecting input files

Before requesting a replication from BPLIM using the original data the researcher must first validate his/her code by successfully submitting the scripts through BPLIM's Replication application. The process involves selecting the main script as well as all the required dependencies created by the researcher. The folder structure used by the researcher is replicated and the BPLIM datasets have to be read from the (read-only) data folder. All intermediary output files must be created during the replication (it is however possible to start from an intermediary output file. In that case, the intermediary file must have been validated in a prior run. BPLIM will then copy the file to the (read-only) data folder.).

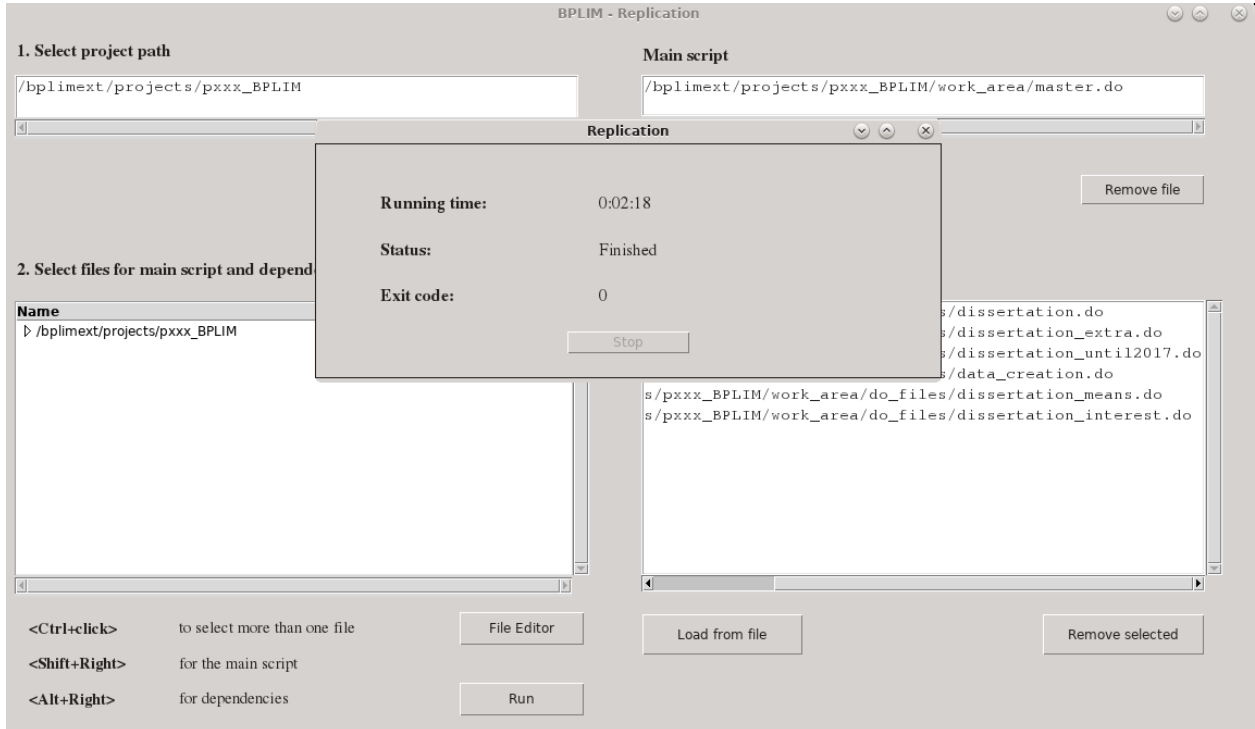


Figure - The BPLIM Replication tool - finished task

The researcher then uses the application to run the code (see Figure 2 for a completed run). The code must run from top-to-bottom and produce no errors. If the run is successful, then implementation on the original data requires only that BPLIM staff changes the relative paths to the data folder and rerun the code. A side advantage of this process is that it automatically produces a replication package for the researcher. Stored in the folder are all replication scripts, the output files, as well as two additional files, one fully characterizing the software environment, and another *json* file containing a listing of all scripts used in the replication. If we add the definition file used to produce the container (or a listing of all packages and respective versions) then we have a full replication package (except for the data).

8.3 Conclusion

Our goal at BPLIM is to make sure that all researchers create their replication packages as an integrated part of their research process. The fact that we are the ones running the code on the original data should be seen as an opportunity to request that researchers make reproducible code while implementing their research.

In the ideal situation that we envision, researchers download a template for the definition file of the Singularity container, customize that template by adding and testing the packages they need, and share with us the definition file. Based on that definition file we build the container for the project and make it available on our external server. The researcher then uses the container to implement the analysis and when he/she is ready to obtain results based on the original data he must first validate the scripts using our application. The researcher can go through this process multiple times and each time a replication package will be created. Apart from the data, all files can be publicly shared, and the replication package created at BPLIM should be easily customized for submission to any data editor.

We are already implementing this solution for all new projects that use confidential data. However, we hope that over time we can convince all researchers at BPLIM to work with Singularity containers and go through the same validation steps that are needed for projects that deal with confidential data.

Projects that are implemented in BPLIM have additional advantages when it comes to reproducibility. First, because all BPLIM datasets are versioned and registered with a Digital Object Identifier (DOI) we are sure that the original data is exactly identified. Second, the computing environment is stable, and the software packages used by researchers are specific to each project. Finally, if external researchers have used BPLIM confidential data then there is an assurance that their code was reproduced at some point.

Ultimately, whether the scientific work is reproducible depends on the researchers. But we hope that integrating reproducibility into the research process with confidential data, provides a way to alleviate the inconvenience of third parties that cannot access the original data and want to verify reproducibility of the results.

Part V

Session 5: Disciplinary support: why is reproducibility not uniformly required across disciplines?

Why do learned societies decide (or not) to implement data (and code) availability policies? What influences the level of enforcement, and the choice of “enforcer” (data editor, administrative staff, referees)? What are reasons NOT to require data sharing or code sharing?

9 Reproducibility in Economics: Status and Update

This article provides a status update on reproducibility in the Economics profession. This includes a particular emphasis on quantitative research. And a chronology of the role the American Economic Association has taken in this progress.

9.1 Background

I am an economist and study poverty and inequality and the role of the social safety net in family and child wellbeing. I want to start by positioning myself and my background a little bit, as it relates to the conversation today. Over the past decade or more, I have engaged in a variety of professional service activities that are linked to reproducibility and open science. First, I am currently a member of the Committee on National Statistics at the National Academies of Science, Engineering and Medicine, where we discuss a range of issues around data, replication and reproducibility. I was a member of the Commission on Evidence Based Policymaking, and there we made recommendations around administrative data and linkages balancing access and privacy. I mention these experiences because in Economics there has been tremendous growth in the use of confidential, proprietary or administrative data and there are significant challenges in open science goals around these data. Second, I spent a decade as a Co-Editor of journals under the American Economic Association, first at AEJ: Economic Policy and subsequently at the American Economic Review, the flagship journal of the association. While a Co-Editor at the American Economic Review, and also a member of the AEA Executive Committee, I Co-Chaired the search committee for an AEA Data Editor. This is the position that Lars Vilbuber holds.

9.2 Main Thoughts

The theme of this session is how differences in support for reproducibility exist across the social sciences. Hence I organize my comments to focus on the status of reproducibility in Economics and what might make it different from other social sciences.

9.2.1 I. The (possibly unique) role of the American Economic Association

The American Economic Association (AEA) has a lot of “market power” because it controls a large share of the high quality journal landscape. The AEA journals include the American Economic Review, the American Economic Review: Insights, and American Economic Journals (there are four of these), [as well as the Journal of Economic Literature and the Journal of Economic Perspectives]. The AEA started with the American Economic Review (AER) and is one of the “top 5” general interest publications in Economics. The AER became so clogged with quality submissions and accepted papers (with 2 year waits between acceptance and publication) that the AEA expanded the number of issues that they produced per year. This led the AER to publish a disproportionate share of the articles in Economics top 5 journals.

Following the success of the AEA journals and the strong budget situation at the association, the AEA decided to expand and add additional journals to its portfolio. We added AER: Insights, featuring short format articles with quick turnaround, in an attempt to compete with publications in Science and PNAS. We also added four top field journals to try to occupy the space just below the top 5, perhaps wresting some of the market from for profit journals such as those published by Elsevier. This became the four American Economic Journals (AEJ) - AEJ: Applied Economics, AEJ: Economic Policy, AEJ: Microeconomics, and AEJ: Macroeconomics.

So if you put this all together, there's a lot of the journal space (and highly ranked journal space) that the American Economic Association runs.

With this “market power” and over period of time the association made a series of moves in the area of open science. This approach came about through both top down and bottom up mechanisms. There was definitely interest among the AEA Executive Committee including the lead editors of the American Economic Review, first Penny Goldberg followed by Esther Duflo. But there was also a perspective that percolated up from the membership. This may have come in part in response to the replication crises in other disciplines. Additionally, many of us engaged in training graduate students desired data for students to get practice and skills.

Overall, I think the AEA views itself as having the potential to take actions that set standards that are adopted more broadly in the profession. (I hope the association takes the same approach to dealing with sexual harassment in Economics.)

9.2.2 II. AEA Actions

So what did the AEA do?

They set up the AEA RCT Registry. As stated on the website: “Randomized Controlled Trials (RCTs) are widely used in various fields of economics and other social sciences. As they become more numerous, a central registry on which trials are on-going or complete (or withdrawn) becomes important for various reasons: as a source of results for meta-analysis; as

a one-stop resource to find out about available survey instruments and data. Because existing registries are not well suited to the need for social sciences, in April 2012, the AEA executive committee decided to establish such a registry for economics and other social sciences.”

Second, and more pertinent to this conversation, the American Economic Review adopted a requirement to post data and code for all accepted papers. It initially was very much on the honor system but it was felt that this was a good first step. There were some staff who worked on confirming that files were uploaded, but not much beyond that.

After some time, there was an evaluation of how this was going. The answer was, not so well. Often times, the data and code would be incomplete. Documentation would be insufficient. It would be difficult for reuse, and it wasn't replicable. There was a recognition that you need skilled leadership and staff for this to work.

This led to the next stage of open science at the AEA. We recognized that we had the budget to do more. The AEA compensates all journal editors a pretty fair wage for the work. And the view was perhaps we should take the next step to hire someone who would work with the journal editors as well as with the AEA Executive Committee to create a more robust system for reproducibility. There was a discussion about whether or not we wanted this individual to be a practicing academic who has skills in this area, or whether we just needed to build up more staffing. We reached out to folks in other disciplines to find out what they were doing. We liked the model at the American Statistical Association who had appointed an academic researcher who also had skills in open science.

We had a search for this new position which we called the AEA Data Editor. From the job ad, the data editor would “Collaborate with journal editors and executive committee; design and oversee implementation of strategy for archiving & curating data, promoting reproducible research.” In short, we needed an architect of the new system as well as a manager and implementer. We were lucky to attract Lars Vilhuber to the position. He is a practicing academic researcher who also has the skills in open science, and could work at the frontier.

9.3 Conclusion

I will conclude by providing some thoughts about where the gaps are and where we need to continue to make progress.

First, we need to continue to revisit the staffing of the Data Editor’s operation. Are the resources sufficient for the needed work?

Second, we need to devise approaches to deal with confidential data. It is increasingly common for AEA journal papers to use proprietary or confidential data. Examples include government administrative data, data from firms, and so on. Researchers typically do not have the ability to post this data. Therefore, replication and reproducibility has to get over a very large hurdle with these data.

Third, we need to build on the success of pre-registration of RCTs. For example, should the journal require pre-registration? Should this be expanded to pre-registration of quasi experimental papers? And relatedly, should the AEA create an RCT Editor to check compliance with pre-registration plan?

Finally, we need do more analysis and landscaping to investigate to where the gaps are in the Economics discipline more broadly, and what role the AEA can take to continue to move us towards further progress in open science.

9.4

Part VI

Session 7: Why can or should research institutions publish replication packages?

This session brings various perspectives together on how research institutions think about publishing replication packages themselves, i.e., not a journal or generalist repository. Panelists come from a university with a specialized, university-centred data repository; from a Federal Reserve Bank with an active researcher community, and from a non-profit (non-academic) research institution. Each faces the requirements of varied internal researchers, external visibility, and differing audiences. The panelists can all speak to how a research institution makes decision about the degree of transparency, and how much of that to do with internal resources.

10 Open Data and Code at the Urban Institute

Abstract: As the leader of the Technology and Data Science team at the Urban Institute, I believe that the field of quantitative social science data analytics needs increased requirements and branding around reproducibility checks, open data, and helpful documentation. Doing so could help to separate high-quality science from less well supported arguments, and pave the way to increasing replication studies in the field. In this piece, I discuss my organization, the Urban Institute, our initiatives, successes, and challenges to date in this area, and my recommendations for next steps.

10.1 Background

The Urban Institute is an organization whose mission is to provide evidence, analysis, and tools to people who make change to ultimately empower communities and improve people's lives. We define "people who make change" broadly as policymakers, government agency employees, advocates, community leaders foundations, corporate leaders, and other similar actors.

Though Urban as an organization has a number of goals, I would categorize our primary drivers as 1) to make impact toward our mission and 2) to fundraise effectively to support that impact and the organizational supports that make it possible. This is important to consider later when I discuss organizational priorities around open data and code.

Similarly, Urban conducts work broadly across many policy areas, however I might summarize them succinctly as 1) conducting policy research and evaluations, 2) providing technical assistance on implementation, 3) producing data and data tools, 4) providing advisory services, 5) convening experts across sectors, and 6) translating research and communicating it to targeted audiences. In support of this work, Urban sometimes posts both the data and code powering the data on its website, Urban.org.

10.2 Main Thoughts

Existing Initiatives

Urban is home to a number of existing initiatives intended to make progress toward more open data and code. The first is Urban’s Data Catalog (<https://datacatalog.urban.org/>), to which all researchers who wish to publish code on Urban’s website must submit their data and document their submissions to a minimum extent. The second is Urban’s central library of quality assurance materials and trainings, which promote open science standards, reproducibility checks, automation in programming, clear organization, and quality documentation throughout. The third is Urban’s automated R (and soon Stata) themes, templates, and guides, which allow researchers to more easily automate the full research pipelines from data collection to publication in R. Urban has processes in place to comply with the requirements of third parties, such as the AEA Data Editor and ICPSR among others, to whom Urban is required to submit or may submit voluntarily. And finally, Urban has a central team that is available to conduct code reviews and reproducibility checks on demand.

Urban continues to make improvements in all of these areas, including adding supporting resources for quality assurance such as an internal code library, providing additional documentation and examples on Github for certain projects, improving our automation of publishing systems to extend to additional content on our website, and revamping and improving our data catalog experience.

Successes

As a result of these efforts, Urban has seen a number of successes that have led to substantial benefits to the organization. For our external users and partners, our publicly available data are now better documented, with a clear license for use, citations are clear and available, and our impact through open data is easier to see and track. I have seen better quality assurance materials and process automation lead to a more streamlined review process that saves time and allows for rapid iteration and even innovation under tight timelines. The processes and systems in place have also allowed for more redundancy and reduced stress on team members when they work well, especially when these efforts reduced in improved documentation, onboarding, communication, and collaboration across teams with diverse skill sets and backgrounds.

Challenges

Despite our efforts and successes, significant challenges remain. Our organization is decentralized and funded by many different parties across government, philanthropy, and the private sector. I have observed that many of these funders in recent years, especially in the philanthropic sector, have shifted their focus away from core research and more toward work that generates impact. I worry that this shift will lead open data and code efforts to be seen increasingly as “optional” when so many other “more impactful” activities are vying for the next marginal funding dollar. In any case, as a result of this landscape, these open practices are

only adopted on a voluntary basis or in certain cases where required by the funder or journal at Urban.

I also observe that researchers and organizational leadership are not directly incentivized, outside of the few funding requirements we do observe (such as the Sloan Foundation, the Arnold Foundation, the National Science Foundation, and others), by existing priorities to tackle these challenges. In the scope of our priority focuses on impact and fundraising, I have observed that while centralized quality assurance and open code are seen as a priority at Urban, they have been at times overwhelmed by even higher priorities, especially in light of my thoughts on funders' changing priorities in this space.

In the meantime, researchers continue to prioritize quality control in their own decentralized individual projects and efforts. However, in my experience the majority are not motivated by quality control arguments to adopt newer open data and code practices, and the strongest motivation remains funder and journal requirements. Most researchers I work with see their work and current processes as high quality, just defined differently across the organization. More importantly, in my view, is that open data and code efforts are often seen as an additional layer of bureaucracy and busy work on top of existing requirements, and thus they are perceived as reducing the agency and academic freedom that many researchers highly value.

10.3 Conclusion

It is hard for me to see openness and reproducibility change without increasing the requirements on researchers and institutions from those funding them and disseminating their work. Ultimately, despite the short-term perceived costs of increased bureaucracy, I believe these requirements will bring larger benefits and are worth considering.

I believe it would be wise for advocates for open data and reproducible research to call for funders and journals to require reproducibility checks at a minimum, and open data where possible as a next step. I would also be in favor of third-party reproducibility checks, and/or marks that certified that a third-party check has been passed and certain materials are available for reproduction.

I believe that these requirements would improve our clients' confidence in the quality of the complying institutions, and clearly help us to differentiate important policy signal from the noise. It would also pave the way toward a future where more replication studies are feasible. Urban currently has systems, processes, and materials in place to comply with these requirements, and the field now has sufficient examples from peer organizations and journals to enable the rapid spread of best practices once requirements are in place.