Safeguarding Research Code & Data: What Economics Journals’ Policies Don’t Get About Preservation

Beyond the Numbers Conference, 2022
St. Louis, MO

Kira Lillard (she/her)
Technical Services Librarian
Federal Reserve Bank of Kansas City

The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.
Introduction

My co-authors: Courtney R. Butler and Brett D. Currier

Started the project in 2019.

Read and analyzed 184 economic journals’ policies on data availability and sharing.
What I want to talk about:

Journal policies that encourage sharing code and data.
Preserving – not just sharing - code and data.
What data and preservation policies look like.
What we do at the Kansas City Fed.
A story about a real person. Esther Duflo.

Esther writes her first economics paper.

She wins the Nobel Prize in 2019.

Every economist in the world wants to read her paper and replicate her findings.
This is not a radical vision.

Librarians, archivists, economists, even economics journals are not solely responsible for safeguarding files.

But they could help.

And not very many of them are.
Why does safeguarding code and data files matter?

Because, you can’t replicate what you can’t find.
“Replicability is essential if empirical findings are to be credible and usable as a starting point for other researchers.”

Ben Bernanke
What economics journals’ policies say.

Economics journals have published policies encouraging researchers to share code and data files for many years, for the purposes of replication.
“Sharing” is not easy.

File sizes
Sensitive information
Contractual terms
Sense of ownership
Fear of plagiarism
Journals’ publishers: A few companies setting policy for many journals.

Elsevier  Springer
Wiley     Emerald
Cambridge Taylor & Francis
Oxford
Policies make sharing sound straightforward.

This journal requires and enables you to share data ...

The journal does not require that research data are shared in a repository...

We require that data be presented in the main manuscript ... or deposited in a public repository whenever possible...
The goal is to be published, even if it requires sharing data.

The American Economic Association implemented its mandatory data availability policy in 2004. By 2018, 74% of economics journals had some form of data policy in place.
A data policy example.

This journal encourages and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles.

https://www.elsevier.com/journals/labour-economics/0927-5371/guide-for-authors
A data policy with preservation features

The journal expects you to archive all the data from which your published results are derived in a public repository. The repository that you choose should offer you guaranteed preservation...

[Link to Wiley's website](https://onlinelibrary.wiley.com/page/journal/14679965/homepage/forauthors.html)
National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation

Storage and Geographic Location

File Fixity and Data Integrity

Information Security

Metadata

File Formats
Beyond bit rot, there’s reference rot.

Referral to a data repository

Assignment of a persistent identifier.

Link to our paper: https://www.kansascityfed.org/research/research-working-papers/safeguarding-research-a-review-of-economics-journals-preservation-policies-for-published-code-and-data-files/
A preservation policy. (Thank you AEA.)

Data and programs should be archived in the AEA Data and Code Repository...If data or programs cannot be published in an openly accessible trusted data repository, authors must commit to preserving data and code for a period of no less than five years...
And publication contingent on sharing code and data.

It is the policy of the American Economic Association to publish papers only if the data and code used in the analysis are clearly and precisely documented and access to the data and code is non-exclusive to the authors.
Percentage of 184 Journals with a Data or Preservation Policy

- Data Policy: 77%
- Preservation Policy: 8%
Percentage of 184 Journals with Preservation Features

- Referral to Repository: 59%
- Persistent Identifier: 47%
- Storage and Geographic Location: 0%
- File Fixity and Data Integrity: 0%
- Information Security: 1%
- Metadata: 15%
- File Formats: 7%
Files don’t last forever. Or even, 20 years.

Replicability of data files means research is credible and usable.

If the file has degraded or the format is obsolete, how useful is it then?
What we do about code and data at the Kansas City Fed

Talk to the economists about replicability and sharing files.
Develop a process to share files.
Create documentation to accompany the files.
Talk to the economists, again.
Nine code and data releases in 2022, so far.

Journals
Press
Colleagues
Nine releases out of a hundred papers in 2022.

And the website is not a repository.
And, we’re not all Esther Duflo.

We should not have to be, though.

KC Fed researchers are doing work on poverty, economic inequity, and gender bias, too.
Thank you.

Kira M. Lillard
kira.lillard@kc.frb.org