









Improving reproducibility in the social sciences: Challenges

Lars Vilhuber

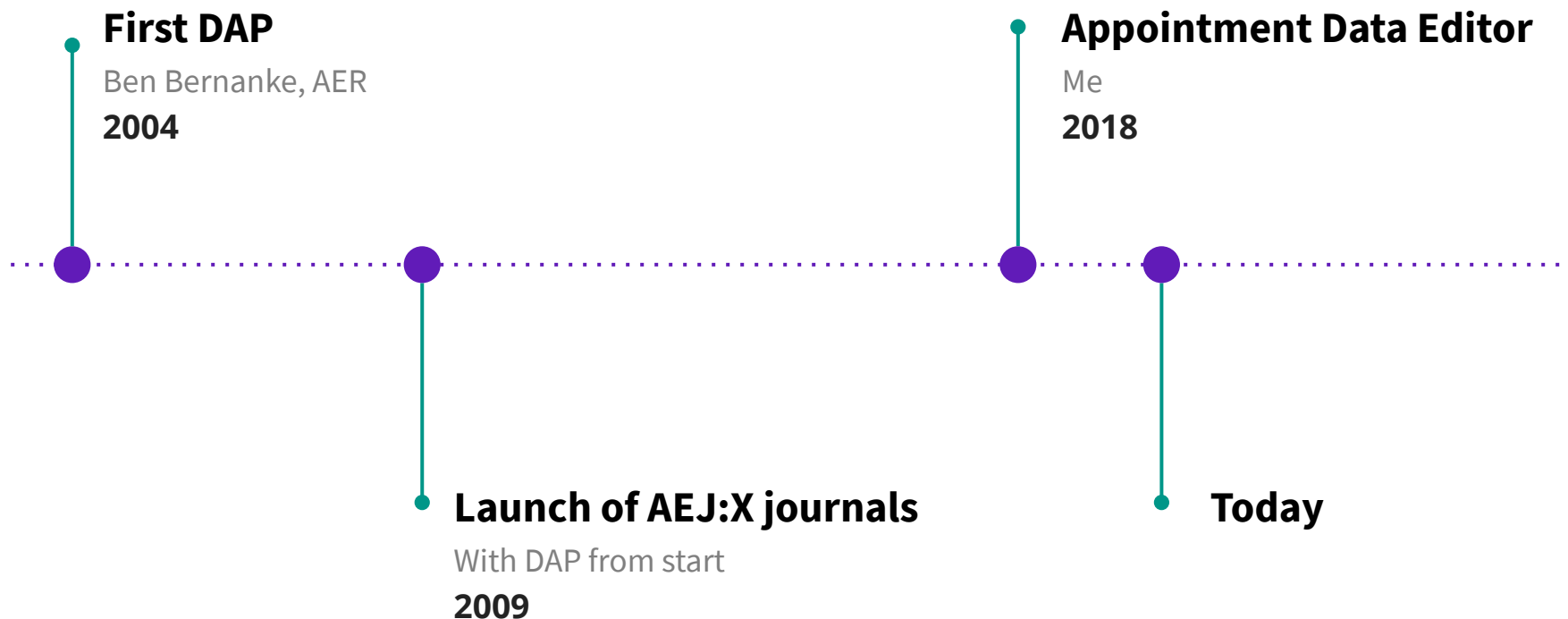
Cornell University / Data Editor, American Economic Association



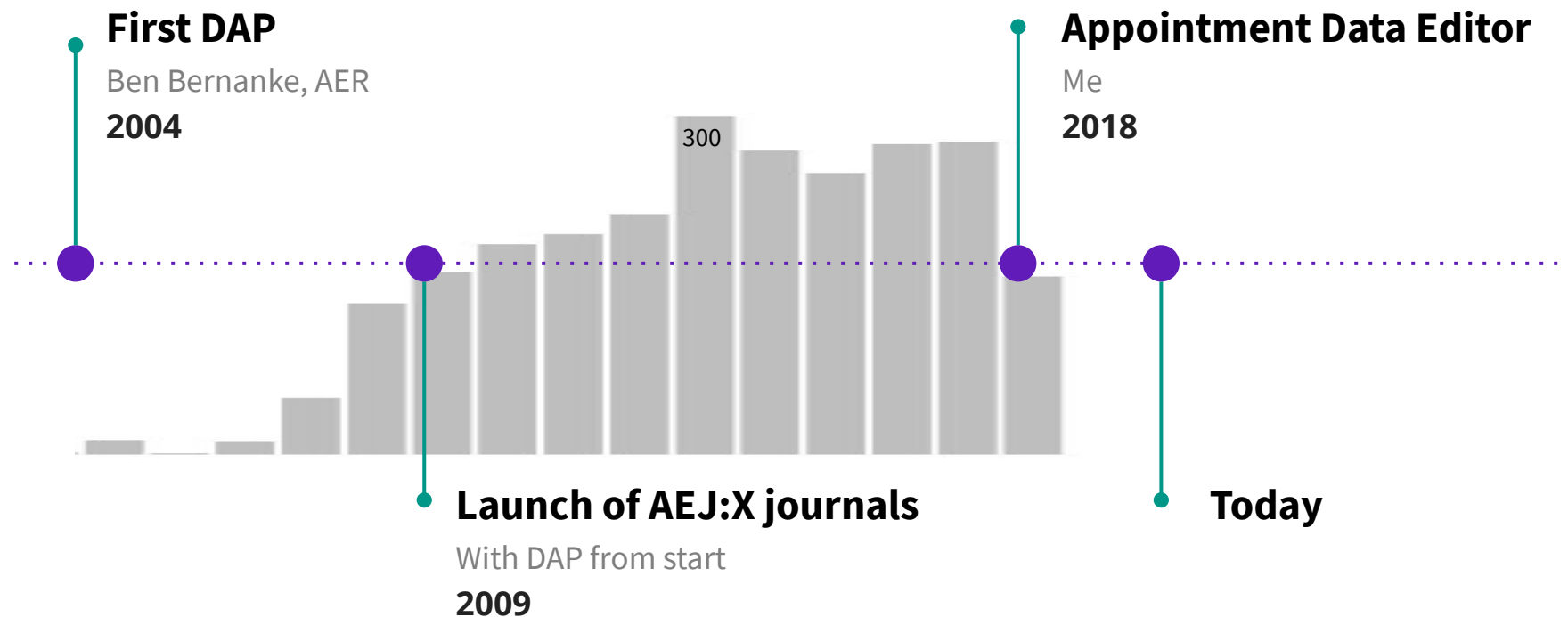
American Economic Association

American Economic Review  <p>The <i>American Economic Review</i> is a general interest economics journal. Established in 1911, the AER is among the nation's oldest and most respected scholarly journals in economics.</p>	American Economic Review: Insights  <p>AER: Insights is designed to be a top-tier, general interest economics journal publishing papers of the same quality and importance as those in the AER, but devoted to publishing papers with important insights that can be conveyed succinctly.</p>
Journal of Economic Literature  <p>The <i>Journal of Economic Literature</i> (JEL), first published in 1969, is designed to help economists keep abreast of and synthesize the vast flow of literature.</p>	Journal of Economic Perspectives  <p>The <i>Journal of Economic Perspectives</i> (JEP) fills the gap between the general interest press and academic economics journals.</p>
American Economic Journal: Applied Economics  <p><i>American Economic Journal: Applied Economics</i> publishes papers covering a range of topics in applied economics, with a focus on empirical microeconomic issues.</p>	American Economic Journal: Economic Policy  <p><i>American Economic Journal: Economic Policy</i> publishes papers covering a range of topics, the common theme being the role of economic policy in economic outcomes.</p>
American Economic Journal: Macroeconomics  <p><i>American Economic Journal: Macroeconomics</i> focuses on studies of aggregate fluctuations and growth, and the role of policy in that context.</p>	American Economic Journal: Microeconomics  <p><i>American Economic Journal: Microeconomics</i> publishes papers focusing on microeconomic theory; industrial organization; and the microeconomic aspects of international trade, political economy, and finance.</p>

Data (and Code) Availability Policy



Data (and Code) Availability Policy



Issues

Generically...

- 100% lack of data citation
- 90% lack of data metadata
- 80% lack of error-free code
- 70% lack of data provenance
- 40% use of restricted data
- 15% replication pkg > 2GB
 - 1% > 20GB


- 73% use Stata
- 22% use Matlab
- 80% secondary use data
- 80% code problems easy
- 90% support reproducibility

... when published at top journals

Advantages

Approximately...

Institutional Enhancements at AEA

- Move to trusted and transparent repository
 - Migrated **~3,000 legacy** replication packages (~500 GB) to openICPSR [[link](#)]
 - Worked with ICPSR to **streamline deposit process** (in progress )
 - **Enhanced metadata** (schema + content) including domain-specific classification system ([JEL](#))
- Implemented a pre-acceptance reproducibility check
 - Verify **computational reproducibility** of every replication package
 - **832** requests for **509** manuscripts
 - Including when **restricted-access data** is involved
 - **36 / 27** successfully accessed via **third-parties**

Disciplinary coordination

Cooperation within economics is critical

- Coordination function amongst editors
- Better preparation by authors
- Training of early-career researchers



Data and Code Guidance by Data Editors

Guidance for authors wishing to create data and code supplements, and for replicators.

Unofficial guidance on various topics by Social Science Data Editors

Guidance on creating replicable data and program archives

This guidance is for the author wanting to create a replication archive.

See [Requested information](#) for the information the Data Editor may request from you, prior to the acceptance of your paper for publication.

Guidance on testing replicability of code

This guidance has two audiences:

- the author wanting to verify whether her code passes muster as a replicable archive
- the replicator wanting to verify the replicability of such an archive

See [Verification guidance](#)

FAQ

See our growing [FAQ](#). If you have questions or answers to add, please notify us by [creating a new issue](#).



A template README for social science replication packages.

The template README provided on this website is in a form that follows best practices as defined by a number of data editors at social science journals.

DOI: [10.5281/zenodo.4319999](https://doi.org/10.5281/zenodo.4319999)

A template README for social science replication packages

The template README provided on this website is in a form that follows best practices as defined by a number of data editors at social science journals. A full list of endorsers is listed in [Endorsers](#).

Versions

The most recent version is available at https://social-science-data-editors.github.io/template_README/. Specific releases can be found at https://github.com/social-science-data-editors/template_README/releases.

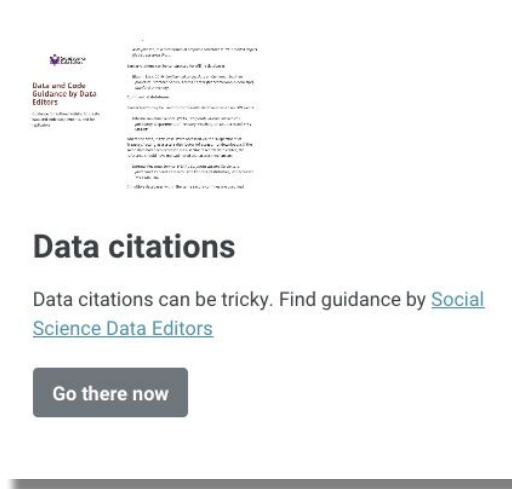
Formats

The template README is available in a variety of formats:

Effecting change

Data citations

- Training
- Monitoring of use
- Verification of correctness
- Collaboration with librarians



The screenshot shows a presentation slide with a purple logo at the top left. The logo is a stylized 'D' with a purple and blue gradient. Below the logo, the text reads 'Data and Code Guidance by Data Editors'. The main content of the slide is a list of bullet points, which are partially obscured by a dark grey rectangular box at the bottom. The text 'Data citations' is visible in a large, bold, black font. Below it, a line of text says 'Data citations can be tricky. Find guidance by [Social Science Data Editors](#)'. At the bottom of the slide, there is a dark grey button with the text 'Go there now' in white.

Data citations

Data citations can be tricky. Find guidance by [Social Science Data Editors](#)

[Go there now](#)

Effecting change

Data provenance

- Clarification
- Verification of correctness
- Repeatability
 - Can others also **find** and **access** the sources?



Data citations

Data citations can be tricky. Find guidance by [Social Science Data Editors](#)

Go there now

Effecting change

Preparing archives

- **Strong** guidance
- README ! (primary doc)



Guidance

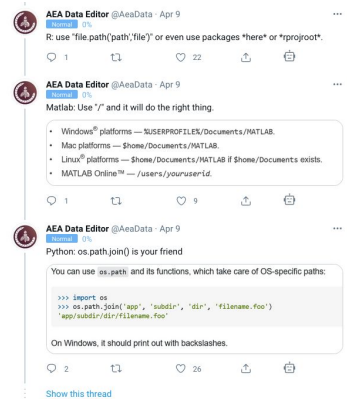
What you need to know to prepare and submit your compliant replication package.

[Read More](#)

Effecting change

Improving reproducibility

- **Strong** guidance
- Examples
- Infrastructure



Effecting change

Greater access to restricted access data

- Work with data providers
 - Private sector
 - Public sector
- Better redistribution
- Better archiving



Effecting change

Education

- Culture of replication
- Improved computing /coding knowledge
- Graduate and undergraduate

Guide for Accelerating Computational Reproducibility in the Social Sciences



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

2021-04-15

[See a full list of contributors](#)

A beta version of the [Social Science Reproduction Platform](#) is now available! [Sign up here](#) if you would like to be part of our beta testing.

Effecting change

Infrastructure

- Easier testing of reproducibility
 - Agnostic to coding styles
- Common to all journals

CODE OCEAN



Singularity



First signs of improvements

- Template README -> improved documentation (at multiple journals)
- Data providers open to collaboration and increasingly streamlining access by editors
- Improved reproducibility? Not yet...
 - The next **509** manuscripts will tell...

Contact

Lars Vilhuber

aeadataeditor.github.io/

