

Without Data, Are We Just Telling Nice Stories?

Josh Nicholson¹

¹Authorea Team

April 28, 2020

At the foundation of research is data. The papers we write and the figures we make revolve around it and it is what we spend countless hours collecting. And yet, most raw data remains absent from major studies (Alsheikh-Ali et al., 2011). This is a problem that has received much attention the past few weeks, with preliminary findings being released from the Cancer Reproducibility Project, a large multi-year effort to see how robust top cancer studies are (201, 2017). Like previous studies in psychology (201, 2015) and cancer (Begley and Ellis, 2012), the findings from the reproducibility project, that a large percentage of findings are irreproducible or at least very difficult to reproduce raise serious questions and doubts about how we conduct and communicate our research.

Authorea was founded to reinvent the research article so that it is data-rich, interactive, transparent, and replicable. Not only did we want to make Authorea a place where researchers could collaborate easier and communicate their results more quickly, we also wanted to make sure that the data behind the study could be easily shared. This is why each article on Authorea is a repository in itself that allows you to host data directly within your article. We enabled integrations with Jupyter notebooks and various data visualization tools not just to make the document more aesthetically pleasing, but to make it easier to analyze each other's work. A quote in *The Atlantic* summarized one problem we're working to fix quite well:

“If people had deposited raw data and full protocols at the time of publication, we wouldn't have to go back to the original authors,” says Iorns. That would make it much easier for scientists to truly check each other's work.- [The Atlantic](#)

We believe that static snapshots of research living in PDFs behind paywalls are inimical to the advancement of research and the findings from the various efforts looking at reproducibility in research support this. Authorea is first and foremost a modern collaborative editor—we want to make it easy to write your work and utilize the power of the web— but we're much more than this, with preprint capabilities (DOIs coming soon), [direct submissions to journals](#), and data hosting, we are working to make research communication more robust on numerous levels. Why should the most important documents in the world be shared and disseminated so poorly? They don't have to be and in fact, we're seeing encouraging signs that the next generation of researchers will do it differently.

The following are just a few **student papers** all utilizing open data sets and analyses on Authorea.

- [Analysis of ground-level ozone formation and its correlation with concentration of other pollutants and weather elements](#)
- [Vision Zero Crash Data Analysis](#)

We hope you'll join us and write your next paper with us. How we make research more robust as a community starts with us as individuals.

References

- Estimating the reproducibility of psychological science. *Science*, 349(6251):aac4716–aac4716, aug 2015. doi: 10.1126/science.aac4716. URL <https://doi.org/10.1126%2Fscience.aac4716>.
- The challenges of replication. *eLife*, 6, jan 2017. doi: 10.7554/elife.23693. URL <https://doi.org/10.7554%2Felife.23693>.
- Alawi A. Alsheikh-Ali, Waqas Qureshi, Mouaz H. Al-Mallah, and John P. A. Ioannidis. Public Availability of Published Research Data in High-Impact Journals. *PLoS ONE*, 6(9):e24357, sep 2011. doi: 10.1371/journal.pone.0024357. URL <https://doi.org/10.1371%2Fjournal.pone.0024357>.
- C. Glenn Begley and Lee M. Ellis. Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391):531–533, mar 2012. doi: 10.1038/483531a. URL <https://doi.org/10.1038%2F483531a>.