# Your Science Is Your (Openly Shared) Data

Your data are no less important than your words.

[Oak Ridge National Laboratory](https://www.flickr.com/photos/oakridgelab/8134349962/in/photolist-doNER9-k4hVBa-6rLmSg-6rQvAm-o5CAVs-6rQvtL-eFkLpu-6tsUAq-fZxQwt-fzWSvQ-6rQu67-6tsUwG-9zmXta-eJNZ9q-fzWRF3-6rQugf-dx9d5a-nkFrFb-fzWSZo-6AvgQd-doNmwz-fzGwDr-doNuFf-k4e4T8-doNueX-efFKR4-dx9cGi-doNuzm-pxXgHn-6rFZp6-e6z2kj-pQd6Nt-e6tpiD-pxZXaU-doNmxT-eFkLH3-doNBKs-e6z2p3-doNBZd-efb554-7RDkKQ-efgM6o-pNmAjY-e6ncFm-e6tppp-efb3Kr-pQsap2-py17U3-efb544-doNBMs)

By Ankur Rashmikant Desai

You just came across an exciting publication in the Journal of Geophysical Research and have a great idea that could be tested—if only you had the data generated by the experiment within. You email the authors, only to learn that the corresponding author has long since retired to the land of dawdling emeriti and the student who collected the data stored them on a crusty, cantankerous magnetic media format for which readers have all but gone extinct. Even when the data are found, information on units, protocols, and column names are garbled, missing, or inaccurate. Woe to the scientist who seeks to reproduce, synthesize, or expand!

Your data are no less important than your words. Get them out there and set them free.

![Every collected data point is valuable and deserves to be saved for posterity—no matter how hard the mosquitoes try to prevent you from getting it. Credit: Connor Klosterman](data:None;base64...)

Sadly, in many cases the opposite is occurring. I run [eddy covariance flux towers](http://fluxnet.fluxdata.org/). Every 50 milliseconds, a wisp of turbulent air exits the forest and washes over my tower, where one sensor feels the pressure of its movement against a puff of sound, and another sensor detects the fade of its infrared light ray source in the presence of various gases. Both sensors’ responses are converted to electrical impulses and sent to the data logger—the keeper of all digits. These digits are radioed back to my lab every half hour, contorted with calculus to make meaning, and placed [online](http://flux.aos.wisc.edu/twiki/bin/view/Main/ChEASData). And while roughly 600 of these towers loom over the globe, only a small fraction of these turbulent eddies, which tell the story of how plants and fungi breathe and live, are preserved and accessible to all. While data, the bedrock of all science, become cheaper to collect, more voluminous, and electronic, our reverence for their sanctity is in worse shape than our current US electoral process. Even the hard-won extraction of the much harder to collect types of data points, such as an airmass isotopic composition from a mass spectrometer, can be lost, if the only published memory of that value is a point on a figure.

As author, you may have noticed AGU journals now require compliance with a data policy in the spirit of [AGU’s position statement on Earth and space science data](https://sciencepolicy.agu.org/files/2013/07/AGU-Data-Position-Statement-Final-2015.pdf). As editor, I have made it my mission to make sure there is teeth to this. I nudge authors to care for their data and data access as carefully as the words they put down on the page to make dreaded Reviewer 2 happy. A statement that data are available in the Acknowledgements from the authors is nice, but I want you to try a little harder. At the least, summarize the key data needed to replicate findings and figures in electronic supplements. Even better, try a public repository.

Set your data free and reap the rewards. In my own work, just the presence of web-accessible raw data from my own lab has significantly increased collaborations, often through chance encounters in late-night Googling by a wayward scholar hungry for just the right observation. Recent [projects](http://pecanproject.github.io/) where I needed to automate assimilation of drivers for an ecosystem model have made it painfully clear to me how important it is that public data be directly downloadable and machine-readable, preferably with [remote subset](http://www.unidata.ucar.edu/software/thredds/current/tds/) capability.

But what about those who appropriate data without attribution? Thieves of bits and bytes? We tend to place high trust in each other, but problems exist and careers can get waylaid. Laying out clear data policies in all your public data, embracing community standards for participation for data synthesis, and calling out the plagiarism that occurs in a manuscript using data without attribution is of course de rigueur. Placing data online does take a leap of faith that the rewards will exceed potential loss of attribution. But more importantly, the act fulfills our fundamental ethical requirement to have our science be open, reproducible, and communicated. In my experience, the bad cases are far fewer than the rewards of open data sharing.

Ok, so you’re convinced—or maybe your program manager or editor is convinced. Make data available. So what do you do? Many disciplines, universities, and federal agencies have started to build repositories, slowly filling caverns of data to mine. The best ones allow for easy uploading and a pathway to making these observations machine-readable, with provenance information and metadata inseparable from the pudding that is your data. Well-known repositories include [DRYAD](http://datadryad.org/), [figshare](https://figshare.com/), [KNB](https://knb.ecoinformatics.org/), [DOE Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics](https://daac.ornl.gov/), [iPlant,](http://www.iplantcollaborative.org/) and [DataOne](https://www.dataone.org/). Some scientists are also resorting to places like GitHub, originally built for software code development, but which is now also a decent home for data, figures, and metadata, even labeled with hashtags (see [#openexperiment](https://github.com/bpbond/cpcrw_incubation), for example). Some disciplines have created their own metadata formats and units, like [Ecological Metadata Language](https://knb.ecoinformatics.org/%22%20%5Cl%20%22external//emlparser/docs/index.html) (EML) in Biogeosciences, or [Climate and Forecast convention](http://cfconventions.org/) NetCDF in Atmospheric Sciences.

Read up on these. Be intrepid and share. Find your repository. Learn about licenses for sharing like [Creative Commons](https://creativecommons.org/). And then you’ll be ready the next time an editor remarks, like in those old Wendy’s ads, “Nice manuscript, but where’s the data?”

—Ankur R. Desai, Editor, Journal of Geophysical Research: Biogeosciences; email: desai@aos.wisc.edu

Citation: Desai, A. R. (2016), , *Eos, ,* doi:10.1029/. Published on 26 May 2016.