



# The impact of sharing supplementary datasets and speaking outside the echo chamber

Alasdair Rae, The University of Sheffield



## Key Points

- **Go the extra mile**  
Any additional information you can share is valuable.
- **Break the echo chamber**  
Sharing your data encourages conversations outside the echo chamber.
- **Sharing data now can save time**  
People don't have to request to see the data.

## Pre-publication

I work in the Department of Urban Studies and Planning at the University of Sheffield.

I've published a few things about flow-mapping. In 2015, I began a piece of analysis looking at census travel to work data in the U.S. I started to get stuck into an origin-destination dataset of commuting flows between about 74,000 census tracts in the U.S.

So, I published the data on Dropbox and I put the working paper on the White Rose Repository at The University of Sheffield. After that, Garrett Dash Nelson, my coauthor (who I only knew from Twitter at the time) took the data and used some algorithmic partitioning tools to break up the commute data into so-called natural communities. In autumn of that year, he blogged the results which he shared with me. That's when we started looking at the data more seriously. Because of the computational complexity, we used Amazon Web Services, which is basically just renting a computer in the Cloud and running the process on that.

So we did that about five times. On the fifth run, we got a result that made the most sense. We submitted a paper to PLOS ONE in July 2016 and it went online in November.

## Post-publication

I'm also an editor of an Open Access journal, *Regional Studies, Regional Science*. I was familiar with the Open Access and sharing infrastructure, so I thought we should do it properly. I spoke to Jez, our Research Data Manager at The University of Sheffield, and said we should get it all on the data portal, ORDA, and repository, WRRO. We also had lots of outputs that couldn't fit in the paper. There were about 35 extra zoomed-in maps that I made which didn't fit, but I thought people would be interested in them. When you do this kind of work, people just want to zoom into their area, so providing this made sense to me.

We also published the raw data and a ReadMe file on figshare.

I've still never met Garrett in person - I suppose it's a story of success for Open Access in some ways but also a triumph of modern communication, too. We've also had a lot of interest post-publication from all over the world. People contact us about everything from Silicon Valley infrastructure to stadium planning to disease mapping. It's been a very interesting experience and figshare has been a big part of it.

Garrett and I also did an AMA (Ask Me Anything) on Reddit recently. This was overwhelmingly positive, but also a validation of the data as lots of people said the regions just 'made sense' to them. Not always, but enough to make us think we got it right.

**“ In the social science world,  
we don't often think of replicability  
and the process of replication.  
But it's good to make sure your  
data is seen and used.”**

## The impact of sharing open data

Had we not been so open with sharing our data, I think we would have had a lot of people emailing us asking for the data and information on how we gathered it. I've had a lot of requests from journalists or people in different parts of the U.S. saying how they felt this data verified what they thought about commuting.

Getting all of our data up there has actually saved a lot of time and effort in the end. The exposure is good, as well: people can spot anything that's wrong and we can easily correct it.

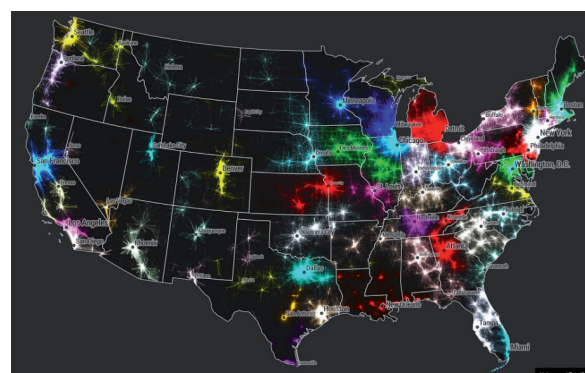
We can measure the exposure with the Altmetrics and stats that are embedded in the dataset's page. As of the end of January, we have over 42,000 views, almost 6,000 downloads, and an Altmetric score of 81, including an article on *Wired*.

If you publish your research and people can't access the data, then, to me, that's partial publishing. In the social science world, we don't often think of replicability and the process of replication. But it's good to make sure your data is seen and used, especially the kind of stuff I do which normally ends up on my hard drive. That's why I started blogging in the first place.

There isn't a downside to making your data available, but it's a bit different when you make all your stuff available. You need to think more carefully about what you're writing and how you write it. The availability of this data should improve practice and scholarly communication. People are really impressed with 10,000 citations on Scholar, but that's all within your peer group - it's just another version of an echo chamber.

I would encourage people to share because you have no idea who is out there. That's the point, I suppose. Otherwise, you'll keep talking to your closed groups that you already know.

If you're doing something similar to the stuff I do, and you do have visual outputs, I would make as many of them available as possible. If it's not that difficult to create additional outputs, go the extra mile to do it, whether it's maps or graphics or tables or whatever. People will be interested.



United States Commutes and Megaregions data for GIS

Get in touch:

**figshare.com**

**info@figshare.com**