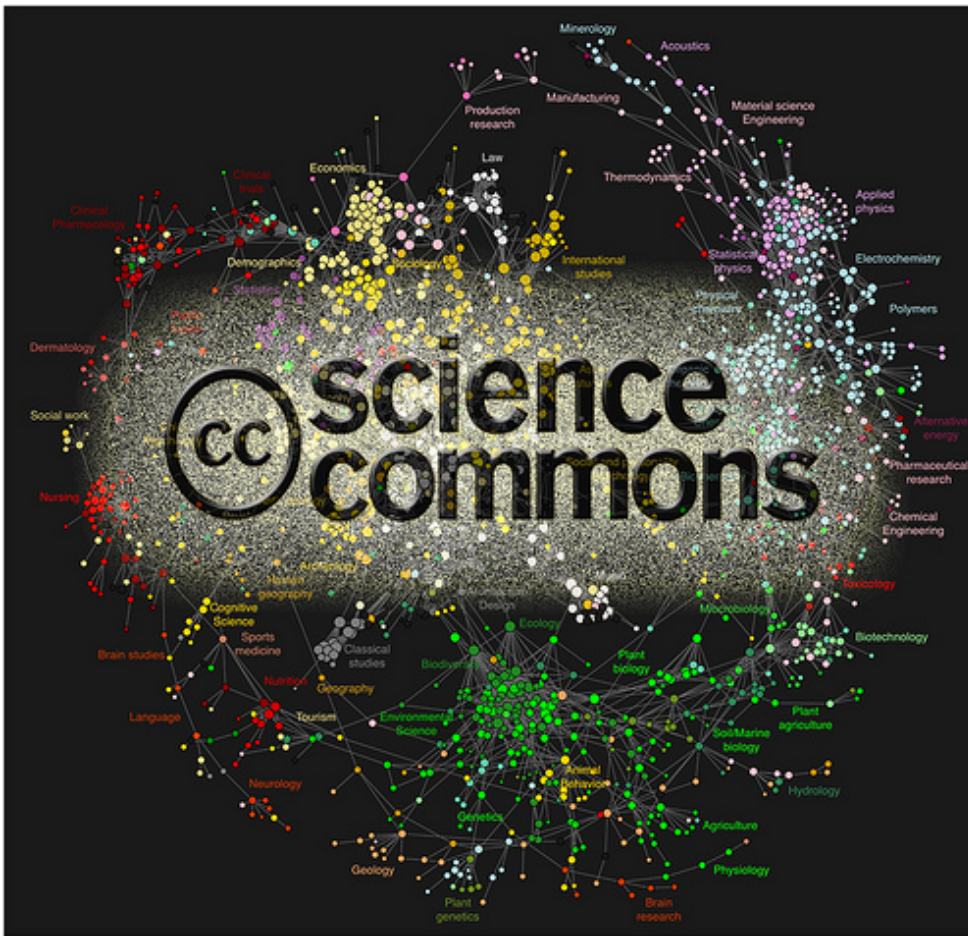


# Replication Data



By: [Duncan Hull](#)

One of the core tenets of scientific research is *replication*. This gets at the *reproducibility* standard of scientific research. Despite calls for more replication in the social sciences, replication studies are still rather rare. In part, this is the product of journal editors' and reviewers' strong preferences for original research. It is also due to scholars not making their data publicly available. Many of my colleagues in academia, especially those who conduct experimental (lab experiments) research, do not typically make their data publicly available, though even here anonymized data should be available.



By: [Seattle Municipal Archives](#)

Replication datasets are not valuable solely for replication studies. In any dataset there are unused variables. A budding scholar or a research-oriented practitioner might be interested in your “leftover” variables or data points. You can’t foresee what others will find interesting.

## What You Can Do

If you have data, share it. Not only is this being generous, but there is some evidence it may even be good for your career (citations, etc.). If you don’t have the capacity to warehouse it yourself, there are archives available for you. A good choice is Gary King’s [Dataverse Network Project](#).

## My Data

In the spirit of replication and extension, I would like to let people know which data sources I have available. If you’d like any of it, shoot me a message and we’ll figure out a way to get it to you.

# Spanish Nationalist Event Data, 1977-1996

First, there is a replication archive of data I used in my dissertation. If you're interested in Spanish nationalist contentious politics – specifically, data on violent and non-violent nationalist protests – check out <http://contentiouspolitics.social-metrics.org/>. This site was set up to make publicly available the data used in my dissertation (2000) and subsequent publications. There you will find background information on the project, codebooks, data, and copies of articles published using the data. You can browse and search the data and view various interactive graphs. The entire dataset is also available for downloading.

## Twitter Data

Twitter data are generally publicly available. However, if you have a pre-defined set of users you can only grab their latest 3,200 tweets, which in some cases is only one year's worth of data. And in other cases, especially if you want to follow a specific hashtag or collect user mentions or retweets, you can only go back one week in time. For this reason, sometimes it can be very helpful if someone else has the historical data you may need. Here are some of the historical data I have, showing the sample of organizations, date range for which data are available, and citations for articles that used the data. If you are interested in it for your own research purposes let me know.

[bibshow file=saxton.bib, format=apa template=av-bibtex-modified]

- *Nonprofit Times 100* organizations – 2009
- *145 advocacy nonprofit organizations* – April 2012
- 38 US community foundations (tweets as well as mentions) – July-August 2011

## Facebook Data

Facebook data for organizations is typically public and can be downloaded via the Facebook Graph API. That said, I have some data available on a sample of large nonprofit organizations.

- *Nonprofit Times 100* organizations – December 2009
- *Nonprofit Times 100* organizations – April-May 2013

## Website Data

Website data. Historical data can often be gathered from the Internet Archive [Wayback Machine](#), but “robots exclusions” and other errors can prevent this. The following datasets are available:

- 117 US community foundations (transparency and accountability data) – fall 2005 (Saxton, Guo, & Brown, 2007; Saxton & Guo, 2011) <sup>[bibtex key=Saxton2007][bibtex key=Saxton2011]</sup>
- 400 random US nonprofit organizations – fall 2007 (Saxton, Guo, & Neely, 2014) <sup>[bibtex key=Saxton2014]</sup>

This is only a partial list of the data I have available. I’ll add to this as more data become cleaned and available.

## References

[/bibshow]