Title:   
**“DataDeps.jl: Automatic Data Setup for Replicable Computational and Data Science Research”**

The open science movement has made great strides in making it possible to reproduce scientific works. The release of data and of source code for research scripts allow simulations and statistical analysis to be repeated.  
However, in practice it is often difficult even given the source code, and documentation to actually replicate someone else’s work. The issue is often not in replicating their results, but rather in getting their research scripts to run at all. These issues come down to problems in matching their computational environment. All non-trivial research code has some dependencies: on software and on data.

Software dependencies are now well managed by modern package managers.  
Data dependencies have not enjoyed the same treatment. Instead dependencies on data are managed either by adhoc solutions, or far more commonly via manual methods.  
This manual data management involves placing instructions in a readme to for example: download a file from a site, extract it either to a particular location, and/or reconfigure the script to reference its location.  
This manual process introduces opportunities for mistakes, and prevents fully automated testing.

This paper highlights the need for fully automated data dependency setup, and discussed out solution: the DataDeps package for the julia programming language.

Julia is a relatively new programming language for technical computing. It has had rapid uptake in the scientific computing and data science communities. It has a strong culture of automated testing, with the ubiquitous use of automatic continuous integration tests (via Travis CI and App Veyor) on all newly created packages/projects.