AMERICAN JOURNAL OF POLITICAL SCIENCE
GUIDELINES FOR PREPARING REPLICATION FILES

Version 1.0, March 25, 2015

William G. Jacoby

The *American Journal of Political Science* requires the authors of all accepted manuscripts to provide replication files before the article enters the production stage of the publication process. The replication files for each article must be made available as a Study located in the *AJPS* Dataverse on the Harvard Dataverse Network (http://thedata.harvard.edu/dvn/dv/ajps). Instructions for getting started on the *AJPS* Dataverse can be found in the "Quick Reference for Uploading Replication Files," available on the *AJPS* website.

This document provides some guidelines, including both general principles and specific requirements, for preparing replication files. The objective is to establish a broad standard for the information that must be made available showing how to reproduce the work that appears in the pages of the *American Journal of Political Science*. This standard should facilitate and encourage active use of the replication files by interested members of the social science research community.

Producing a single set of instructions that would pertain to all articles appearing in the *AJPS* likely would be impossible. So small exceptions may be necessary in order to accommodate some individual articles. But, the following guidelines should sufficiently address the vast majority of situations that arise.

Note also that the following guidelines concern the replication materials for quantitative analyses. The same general principles should apply to qualitative analyses, but the specific requirements probably will differ. Therefore, guidelines for qualitative data are currently under development and will be added to this document in the near future. Until then, authors of articles that rely on qualitative data should contact the *AJPS* Editor for further instructions.

General Principles:

The purpose of replication materials is to enable interested researchers to reproduce all of the analyses that are reported and presented in articles that appear in the *American Journal of Political Science*. Authors should always consider this basic objective when preparing replication materials. Generally speaking, more information is better than less.

All analyses reported in *AJPS* articles are open and available to the scientific research community. Authors are not permitted to "embargo" or withhold information that has been used to perform an analysis featured in an *AJPS* article. Instead, authors must provide all information that is required to reproduce any analytic result that is reported in their article. This material includes (but is not limited to) analysis data, code for running relevant software, and source files for software that is not publicly available (e.g., Stata .ado files or R packages written by the author).

Authors do not need to provide any additional information or data, beyond what is necessary to reproduce the analyses reported in the *AJPS* article. So, for example, the analysis dataset may contain variables that are neither employed in any models, nor used to construct variables that are employed in reported models. If so, then these variables need not be included in the replication materials for the article.

Specific Instructions and Recommendations:

In most cases, the replication materials will include: a Readme file, information about the data source(s); the analysis dataset; and code for running relevant software. Each of these elements will be discussed in greater detail, below. Optionally, authors may include extensions to the analyses reported in the article. Note that the replication materials are expected to cover all analyses presented in the main article and in the Supporting Information for the article.

Readme File:

Every Study on the *AJPS* Dataverse should include a plain-text file named "readme.txt". This file provides the names of all other files contained in the study, along with a brief description of each one. For relatively small Studies, this information can be presented as a simple list. Larger Studies should group similar files under subheadings, such as "Data files" and "Stata .do files", or "Files to Reproduce Table 1", etc.

Analysis Dataset(s):

Every Study must include one or more files containing the data required to reproduce all tables, figures, and other analytic results reported in the *AJPS* article and its Supporting Information. Authors can choose their own data storage formats (e.g. rectangular text files; comma-delimited files; Stata .dta files; data objects within an R package; SAS files; SPSS files, etc.), as long as the files are readily accessible to researchers in the social science community. Files stored in arcane or proprietary formats generally are not acceptable.

Data should be arranged within each file to facilitate understanding of the contents. If possible, variables should be given meaningful names. And, a unique case identifier variable always should be included. If the data are extracted from another source (as often will be the case) then the case identifier should link the observation in the analysis dataset to its information in the original source.

In analyses based upon highly data-intensive procedures (e.g., Bayesian simulations, bootstrap resampling, etc.) it is not necessary to provide the full contents of each replicated dataset. However, the full set of relevant results (e.g., the simulated parameter values in MCMC estimation, the bootstrap replications of a sample statistic, etc.) should be provided in a coherent data file. And in such cases, providing software command files (see next subsection) to reproduce the entire data-intensive analysis is particularly important.

Each file containing an analysis dataset must be accompanied by a PDF file containing a codebook for the dataset. The codebook always should contain variable definition information for all variables used in the analysis. If the dataset is stored in a proprietary format (e.g., a Stata .dta file) then the codebook must include variable names. If the data are stored in a text file, then formatting information must be provided.

In some rare cases, the analyses in an *AJPS* article may be based upon restricted data that cannot be posted in a publicly-accessible location. Any such exceptions to the general *AJPS* policy regarding replication materials must receive explicit prior approval from the Editor. Once this permission is

received, the analysis dataset need not be posted to the *AJPS* Dataverse. But, the author still must provide instructions that interested researchers can use to access the data (see the section on "Information to Reproduce the Analysis Dataset" below), as well as formatting and variable definition information for the data that are analyzed in the *AJPS* article.

Software Commands:

Every study must include one or more files containing the software commands that can be applied to the analysis dataset in order to reproduce all tables, figures, and other analysic results presented in the *AJPS* article. Typically, these will be plain text files. The exact format of the file contents depends upon the software used to carry out the original analysis. Authors can provide Stata .do files, R command scripts, or text files appropriate for submission to other software systems and environments.

Regardless of the format used for the command files, comment statements should be used extensively throughout the files to explain the steps of the analysis. Authors can assume that users are familiar with the software system used for the analysis (i.e., there is no need to explain how specific commands function). But, authors should explain how the various commands produce results that are relevant for the analyses reported in the article (e.g., "The following commands recode variables X and Y in preparation for the logistic regression model"; "The following commands create Figure 1 in the article"; etc.).

In some cases, conducting the analyses in an *AJPS* article may require software tools that are not readily available to the research community. Examples include (but are not limited to) Stata .ado files or R packages written by, or specially available to, the author. Any such software resources required to replicate an analysis from the *AJPS* article must be included in the Study for the article, along with relevant documentation and instructions for installing (if necessary) and using them.

Authors always should provide clear and specific information about the version of the software system used to conduct the analyses reported in their *AJPS* article. This requirement is critically important because algorithms, procedures, and functions can (and do) change across software versions! For example, an R command file might begin with the comment statement, "The following analyses were carried out using R version 3.1.0" or the Readme.txt file for the Study could contain the following statement: "All data analyses in this article were carried out using Stata/MP 13.1 for Windows (64-bit x86-64)."

As mentioned earlier, software command files are particularly important for data-intensive analyses in which the "intermediate" datasets used to obtain the final results (e.g., MCMC simulations or bootstrap replications of the original data) are not, themselves, included among the Study's replication files. In such cases, users literally would be recreating the original analysis. Thus, the command file needs to provide especially clear instructions for doing so. On a related point, any commands that generate random numbers (e.g., for Monte Carlo simulations, bootstrap resampling, jittering points in a graphical display etc.) should include a seed value in order to insure consistent results.

Authors can provide either a single command file that covers all analyses reported in the *AJPS* article, or separate files for the specific analyses contained within the article. In the former case, comment statements should be used within the command file to distinguish the commands used for different figures, tables , or other analytic results. The process is best served when authors use meaningful, rather than "generic" file names. For example, files named "Stata commands for performing logistic regressions.do" or "R functions to reproduce Figure 1.R" are better than "Commands.do" or "R_scripts.R".

Information to Reconstruct the Analysis Dataset:

Every Study must contain complete information for constructing the Analysis Dataset(s) from the original sources of the data. The exact materials for doing so will depend upon the nature and sources of the analysis data. But regardless of the specific details, interested researchers always must be able to follow the author's instructions in order to reproduce the precise data values used for any analyses reported in the *AJPS* article.

The Analysis Dataset often is created by extracting variables and observations from another, larger, dataset such as an entry in the American National Election Study series, The Comparative Study of Electoral Systems, the International Correlates of War project, or the General Social Survey. In such cases, the author must provide a software command file for doing so. Although the format will vary depending on the software that the author uses, the file always should contain commands for selecting the relevant variables, extracting subsets of observations if necessary, performing any data transformations that are carried out prior to the analysis itself, and assigning missing values. Again, comment statements should be used extensively throughout the file to explain the commands that are used. As stated earlier, there always should be a case identification variable that links observations in the Analysis Dataset to their original records in the source dataset. And, it is important to identify the specific version of the source dataset and the date that it is accessed in order to construct the Analysis Dataset.

The Analysis Dataset sometimes is created by merging information extracted from several other sources. For example, an analysis of the American states may use information obtained from both the states themselves and from the federal government. In such cases, the author must provide the relevant software commands for extracting the data from the separate sources, and for merging the separate subsets of data into the overall Analysis Dataset. As always, comment statements should be used extensively to explain the procedures.

Complete reference information must be provided for all source datasets used to construct the Analysis Dataset. Some useful information about the practice of data citation can be found on the website of the Inter-university Consortium for Political and Social Research (ICPSR):

http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/citations.html

Restricted Access Source Datasets

There are two situations in which an author of an *AJPS* article can request, and may be granted, an exemption from the general requirement of providing the source data for the Analysis Dataset. First, the author may want to prevent outside access to the source dataset in order to further his or her own research. As explained earlier, the data used in the *AJPS* article cannot be "embargoed" in this manner. However, the author can request an embargo for any additional information that is contained in the source dataset(s). Any such restrictions will be imposed for a limited amount of time only, and the Study materials will include a statement specifying when the source data will be made available to the general research community. The *AJPS* Editor explicitly must give permission for any data embargo of this type. Furthermore, the author still must make the source data available to the *AJPS* Editorial Staff and to the *AJPS*' contractor tasked with verifying the content of replication materials (currently, the Archive Staff at the Odum Institute for Research in Social Science, University of North Carolina at Chapel Hill). The AJPS Editorial Staff and the Archive Staff at the Odum Institute guarantee that the embargoed data will be used strictly to verify the integrity of the replication materials, and will not be retained after that has been established.

The second situation involves a restricted-access source dataset. In such cases, the *AJPS* Editor may allow an exception to the general policy of full data access. Note, however, that "blanket" denial of access to the Analysis Dataset used in an *AJPS* article will not be permitted. The holders of the restricted-access data also must provide the *AJPS* Editor with (1) a statement that access will be granted to the data for qualified researchers; (2) an explanation of the conditions under which the data will be made accessible; and (3) the qualifications that a researcher must present in order to be granted access to the data.

If the *AJPS* Editor does permit restricted access to the Analysis Dataset or the source data, then the author must specify the procedures through which an interested researcher can apply for access to the Analysis Dataset for replication purposes (including the construction of the Analysis Dataset from the original source dataset) from the holders of the source data. The author also must provide software command files for carrying out the analyses and constructing the Analysis Dataset. The objective here is to provide researchers with all possible information about the data analysis, even if the data, themselves, are not immediately available in the Study on the *AJPS* Dataverse.

Conclusion

The preceding guidelines describe the minimum requirements for a Study on the *AJPS* Dataverse. They implement the principles of data access and research transparency (DART) to which the American Journal of Political Science is a founding signatory. But, authors certainly are not limited to providing the previously-described files. In fact, they are encouraged to provide as much information as possible. Additional contents of the Study on the *AJPS* Dataverse might include supplemental reports, pre-analysis plans, additional data, and extensions of the analysis beyond those reported in the *AJPS* article and Supporting Information. Authors are encouraged to keep in mind that scientific research is an ongoing stream. Hopefully, the research reported in the pages of the *American Journal of Political Science* will facilitate and encourage further efforts to establish powerful theories of political and social phenomena. Any materials that facilitate this process are welcome elements of a Study on the *AJPS* Dataverse!