AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

AMERICAN JOURNAL OF POLITICAL SCIENCE
QUALITATIVE DATA VERIFICATION CHECKLIST

Version 1.0, March 11, 2016

The following checklist can assist authors in preparing qualitative data they will submit to the *AJPS* Dataverse after their manuscript has been accepted (and before the production stage of the publication process) to facilitate evaluation of the empirical claims they make in their article. The list is not exhaustive and not all aspects will be applicable to all articles or all types of qualitative data. However, by addressing all relevant aspects listed below, most authors will be able to provide sufficient information to make their data accessible, and render the analytical processes they used explicit and verifiable. Authors should also consult and follow the instructions provided in the *AJPS Guidelines for Preparing Replication Files*.

Supporting Documentation and Information about Procedures:

- ☐ Includes a README file (.txt format) containing the names of all files with a brief description of each one; a brief discussion of how the data were collected (when and where they were obtained, data collection techniques used, personnel involved); a brief description of the data-collection instruments used (e.g., survey questionnaires, interview protocols); a brief discussion of the text collection procedures employed; a full list of the files included in the analysis dataset with a brief description of each one; and, where annotation is used, the rules the author followed in deciding which material to annotate (e.g. centrality, contestation).

- ☐ Where applicable, includes a codebook (.txt format) with code or variable definitions and any value labels used; a document term matrix representation of the text collection and a description of any text processing steps undertaken prior to analysis (regardless of whether the full texts are also provided) or a list of coding categories, final coding trees, and outputs.

- ☐ Provides clear information about any proprietary data formats (particularly applicable if R or similar statistical packages, in the case of matrix-type data, or CAQDAS were used), including the particular software version used.

- ☐ If necessary, gives instructions regarding access conditions the author wishes to apply to some or all of the publicly available version of the analysis data, providing the relevant justification for any restrictions and exceptions (these instructions, like all other documentation, will be made public).

Analysis Data:

- ☐ Includes files containing the data that were analyzed in order to make the central and contested empirical claims in the article. Data must be provided in readily accessible file formats (delimited values, .txt, PDF, or JPEG are preferred).

For data organized in matrix form (approached as an aggregate body of information and analyzed holistically):

- ☐ Where relevant, includes software command files for reconstructing the analysis dataset or code for the calibration of sets, construction of truth tables, and so on.

- ☐ When the analysis dataset is extracted from an existing database, provides the technical means to derive the analysis dataset (for example, unique case identifiers to the original data sources).

- ☐ Includes all of the materials (commands, information about software packages used, etc.) necessary to reproduce the analyses, tables, and figures reported in the article.

- ☐ Includes instructions and commands for installing any non-standard software macros or packages.

For granular qualitative data (i.e., data drawn from individual sources or small groups of individual sources considered separately to draw conclusions), each file must include either:

- ☐ A verbatim or redacted fair-use excerpt of about 100 words, providing sufficient context to make the source understandable and verifiable.

- ☐ In circumstances where no excerpt can be provided due to human subjects, copyright or logistical constraints, an explanation of those restrictions and a summary or paraphrase of the relevant text.

☐ File names must follow a consistent, human-readable pattern. We recommend using the relevant elements from the following naming convention:

<author's last name>_<brief description>_<date of original publication/creation>_<identification label>

An “identification label” is most appropriate for individual sources. Identification labels are described below, in the last point under “Analysis Procedures and Annotations for Transparent Inference.”

Examples of this file naming convention include:

Brown_TruthTable1_2016
Brown_NVivoCodingTree_2016
Brown_Interview1_1995_1a
Brown_NewspaperWSJ5_2000_14b
Brown_Huntington_1993_54a

☐ Includes transcripts of relevant passages from handwritten or audiovisual materials.

☐ Includes translations of relevant passages (e.g., 100-word excerpts) of any non-English materials, including the source of the translation.

- ☐ Includes complete bibliographical and location information for the source(s) used to produce the central and controversial claims in the article. This could be achieved by filling out the relevant fields in a Transparency Appendix (TRAX) template. Templates can be requested from the Qualitative Data Repository (QDR) at Syracuse University.

- ☐ In addition to the other bibliographical information, includes URI's (typically URL or DOI) for any electronically available sources

Analysis Procedures and Annotations for Transparent Inference:

- ☐ Supplies annotations describing the micro-connection between the data provided and the central, controversial claims made in the article when this connection is not already elucidated in the main text of the article.

- ☐ Ensures that data files and annotations are linked or organized in a way that makes the links between them clear. We recommend using an annotation tool (e.g., Word "comments" for a file in .docx format) to include unique identification labels for each location in the article text where data are cited to support a claim. Use numbers in sequence starting with "1". Number the supporting information (i.e., a more complete citation, an excerpt, an annotation, etc.) in the supplemental materials in the same way. If more than one source is used per claim, use "a", "b", "c", and so on after the sequential number. Using a Transparency Appendix (TRAX) template (available at QDR) can facilitate the process of clarifying links in this way.