Data Management Plan

Roles and responsibilities
Dr. Carwil Bjork-James, the PI, will be the primary person responsible for data collection, storage, and analysis. He will train and supervise two graduate student researchers and undergraduate research assistants in data collection and management. Undergraduate researchers will be trained as part of Vanderbilt's Research on Conflict and Collective Action Lab course. The database, supporting documents and media are stored in online spaces. Nonpublic documents—the audio and transcripts produced in interviews and researcher fieldnotes—will be held by researchers and handled as described below. At the end of the research project, retained data will deposited and managed via the Qualitative Data Repository at Syracuse University (https://qdr.syr.edu).

Expected data
This project will generate data in three formats:
1. **A database** of information related to deaths in political conflict, which will be stored and accessed as follows:
    a. As a *Google Docs spreadsheet* for simplified and validated data entry.
    b. As *R dataframes*, with revisions archived in .csv files, and public release versions archived to a GitHub repository using datastorr.
    c. Our suite of *R scripts for analysis* will be archived to, and documented in an R package archived to separate GitHub repository. Semantic versioning standards will be used, separately numbering releases of the dataset and the tools for analyzing it.
    d. As an *R Shiny-based website* for accessing, filtering, summarizing, and visualizing the data. The PI will work with a student programmer to design and code an interactive interface for exploring the dataset.

    This database is currently stored as a Google Docs spreadsheet, and analyzed using a combination of summary tables and charts within that spreadsheet, as well as R scripts that scrape the main page of entries using the R package googlesheets. In the interest of flexibility and open access to the data, we will migrate all analysis to R scripts, and use an R Shiny app to allow users to summarize or visualize the entire dataset, or only deaths that meet particular criteria.
2. **Texts**, including:
    a. Narrative and summary texts describing the events in which people died, the larger context, and any uncertainties or disputes in the source data.
    b. Digitized copies of news reports, human rights reporting, and source documents that document the deaths.
    c. Transcripts of interviews.
    d. Field notes on researcher observation of public events and social movement campaigns.
    e. The codebook for the database

    Narrative and summary texts will be written in Wikipedia-compatible Wikitext, which allows for full documentation of sources for each fact. Digitized source documents (2b), interview transcripts (2c), the codebook (2e) and other text documents will be archived in text and Markdown formats. All nonconfidential documents will be shared among researchers via Dropbox, indexed with Zotero, and archived along with the database to the QDR.
3. **Audiovisual files**
    a. Photos of archival documents. These photos will be stored in their original format (.jpg and/or .raw) but later archived as .tiff files.

b. Recordings of interviews using digital recorders or video cameras. Audio files will be saved as .wav to maintain high quality. Video files will be saved as MPEG-4_AVC files to maintain quality and to facilitate easy playback.

Data retention, storage, dissemination, and preservation of access

The database (1b, 1c), alongside supporting files (2a, 2b, 2e) will be archived along with the database to GitHub, and made available within R via datastorr. Through datastorr, researchers seeking to replicate our analysis will have one-line access to importing the data into R. Release versions will also be permanently archived and stored in the Qualitative Data Repository, where they will have metadata, documentation, and a permanent doi identifiers. Public releases will include only those variables that we have reviewed for quality and completeness. Once it is developed, scholars and the general public will access our database using the visualization/exploration tool (1d) on our website.

Portions of the summary texts (2a) that are compliant with Wikipedia policies (i.e., notability, reliance on publicly available secondary sources, and no original research) will be published directly to Wikipedia, something we have pilot-tested for the 1988 Villa Tunari massacre. Subject to advice on copyright restrictions, we will also share source documents or relevant quotations (2b, 3a). The project will maintain a website linking to such data, sharing research findings, and introducing and sharing selected source documents found during research. Other supporting documents will be made available to other researchers upon request, subject to the limitations of privacy and confidentiality described below. Audio files and field notes will be kept in password-protected files on the PI's computer.

Interviews (2c, 3b) will be handled according to standards of oral history: Interview subjects will be asked whether the recordings of their interviews may be shared publicly or not, and if they will be identified by name, by organization, or anonymously in public releases. Subjects who consent to being identified publicly may also be photographed, or have all or parts of their interviews video-recorded. At the end of every interview, researchers will remind subjects of these options and their ability to embargo release of their interviews to a future date. Per oral history standards, interviewees will review and approve the transcripts of their interviews for accuracy and for public release. Researchers will respect interviewee requests to delete or embargo all or part of their interview segments. When interviews are approved for public release, they will be added to the collection of publicly available supporting documents. Any material which subjects have not requested to be removed or kept confidential will be archived for the use of future researchers (as de-identified transcripts if anonymous, or as transcripts and edited audio if not) via the Qualitative Data Repository. If the subject requests confidentiality, the researchers will destroy the original audio and visual recordings of interviews as promised on the consent form.

The PI has chosen to archive disseminate data via the Qualitative Data Repository to guarantee free access to future researchers and reliable institutional management of all materials. Qualitative Data Repository staff follows high standards for storage, maintenance, and accessibility of research materials; it is easily accessible by researchers around the world, has a dedicated management staff, and accepts audiovisual files. Our data will also be indexed in the American Anthropological Association Data Registry Wiki, and cited via its doi in scholarly and online publications to facilitate data sharing, citation, and further discovery.

Field notes (2d) will be archived by each researcher as follows: The PI and researchers will not make field notes available at this time per the American Anthropological Association's Statement on the Confidentiality of Field Notes (2003). Due to the potentially politically sensitive nature of this research, we plan to embargo this material until the end of the interviewees' careers in public service, community leadership, or advocacy. Ethnographers have an ethical observation to preserve their field notes, and this plan assigns that responsibility individually to the ethnographers, who will maintain control over their notes after the completion of the project.