

Data Management Plan

Expected Data

For this study we will collect data from three surveys and one wave of qualitative interview data from an international military base site visit. The survey will be nationally representative surveys of adults (18+) in Djibouti, Kenya, and Cambodia—though with quotas to ensure sufficient representation from areas around military bases. These surveys will be distributed by a reputable survey sampling firm, who will recruit participants. All individual survey responses will be de-identified. The datasets will include basic demographic information (e.g. race, age, gender, etc.), responses to questions about the participants' political and social attitudes and behavior, and data about their views of the United States, China, and those two countries having a military presence in their country.

For these surveys, no information capable of uniquely identifying the respondent will be collected as part of data collection. Should local sample vendors include any unique identifiers (such as respondent IDs), these will be removed from any dataset. We plan to share individual-level respondent data, so that other researchers can conduct their own original analyses and verify our analyses.

The qualitative interview data will be gathered from stakeholders during a site visit to a Djibouti. The primary purpose of these interviews will be to gather information that can contextualize the quantitative survey findings. Stakeholders will include Djiboutian and U.S. government officials and military personnel stationed on foreign bases on Djibouti. An interview protocol will be developed for site visit interviews and responses will be recorded by hand. Handwritten notes will then be transcribed into an electronic format for analysis. A summative report will be produced based on the interview data.

Management and Maintenance of data

Data will be stored in a number of places, each requiring different security procedures. Survey data will be collected by the survey firm GeoPoll, and deidentified raw response data will be retained on GeoPoll's secure servers. No one except the Co-PIs will be given access to the data. Data will also be downloaded, cleaned, and formatted. These datasets will have no uniquely identifying information included. Local copies of these datasets will be maintained on computers used by the Co-PIs.

Interview notes from the site visit will be recorded by hand and then transcribed into an electronic format. The PI will conduct the transcription. The original notes will be stored in a locked filing cabinet and maintained for three years, at which time they will be destroyed and only the electronic versions retained. A unique identifier will be given to each interview and no personally identifying information pertaining to the interviewees will be recorded. The unique identifier will be used for the purpose of distinguishing data gathered from military personnel/stakeholders compared to data gathered from host country stakeholders. All computers will be password protected and maintain up-to-date antiviral software to avoid security breaches related to malware.

Factors possibly impacting ability to manage data

The primary issues we face in managing and distributing our data are related to confidentiality. We are collecting public opinion data, which may include opinions and experiences that respondents would prefer not become publicly known. We have protocols in place, including limiting access to the survey data, and using password- and antiviral software-protected laptops to prevent any data breaches. Respondents will

be informed of the (very minimal) risk of identification due to security breaches as part of the process of gaining informed consent.

Mechanisms for data sharing and Metadata

We will use Harvard Dataverse to distribute data to the public. Data will be made publicly available at the conclusion of the project's funding period. Harvard Dataverse stores a great deal of metadata, such as collection dates, associated publications, sample vendors, etc. We plan to take full advantage of these capabilities and enter any and all metadata we can, so that other researchers can have access to it. We will work with librarians at University of Miami to ensure we have appropriate metadata available for each dataset. Additionally, we will develop a codebook for both datasets, with variable descriptions, coding rules, and any relevant notes that will be stored along with the dataset. The qualitative data (interview notes from site visit) will be disseminated in the form of a summative report. Given the small sample size and need to maintain confidentiality, the raw interviewer notes will not be made publicly available.

Other details

Period of data retention. All quantitative data will be made publicly available via Harvard Dataverse as soon as the paper using the datasets is accepted for publication, or when the grant period ends, whichever is sooner.

Data formats and dissemination. All data will be stored either in secure cloud servers (including but not limited to those maintained by GeoPoll, Box, and Dropbox) or on password- and antiviral software-protected computers maintained by the project team. Uniquely identified data will only be accessible by the Co-PIs. Quantitative data will be made publicly available (after an embargo period) via Harvard Dataverse. Datasets published through Harvard Dataverse will be released under a modified Creative Commons license. Additionally, descriptive metadata records and permanent DOIs will be created for each dataset to ensure proper citation and permanent retrieval of the materials.

Data storage and preservation of access. All electronic materials will be stored in a secure cloud storage location (password-protected Dropbox or Box account) on a password-protected computer. The PI will work with their institution to ensure continued access and storage of these data meet the University's data management requirements. Research data generated by the project will be openly shared via Harvard Dataverse. Additionally, descriptive metadata records and permanent DOIs will be created for each dataset to ensure proper citation and permanent retrieval of the materials.