

## Plan Overview

---

*A Data Management Plan created using DMPTool-Stage*

**Title:** Ethnic Diversity and Public Goods Provision Across Latin America

**Creator:** Christopher Hale

**Affiliation:** University of Alabama (ua.edu)

**Principal Investigator:** Christopher Hale

**Data Manager:** Christopher Hale

**Funder:** National Science Foundation (nsf.gov)

**Funding opportunity number:** 59262

**Template:** NSF-SBE: Social, Behavioral, Economic Sciences

### **Project abstract:**

This research project seeks to address two related, but distinct, research questions: 1) Is ethnic heterogeneity negatively associated with public goods provision across Latin America? As my preliminary attendant statistical analyses have revealed there is no negative relationship, I subsequently ask: 2) Why is ethnic diversity not negatively associated with public goods provision across the region? I use a multi-method research design exploiting variation across time and space in Latin American public goods provision utilizing: A) an originally-constructed panel data set consisting of 17 Latin American countries from 1990 to 2014 (425 observations) to assess several potential explanatory factors' influence on public goods provision in Latin America as measured by the infant mortality rate; B) a second statistical analysis examining individual-level political engagement with survey data collected by the Latin American Public Opinion Project (LAPO) at Vanderbilt University for a total of over 130,000 individual observations spanning the time period from 2004 to 2014 across 17 Latin American countries; C) qualitative field research of nearly 6 months located in Cauca (Colombia), Oaxaca (Mexico), and Sololá (Guatemala). In my qualitative analysis, the mechanisms linking ethnic heterogeneity to public goods provision will be assessed within each of these three regions. I will triangulate the strengths of each methodological approach to test several theories that potentially explain public goods provision across the region. These will be assessed against my own working hypothesis that where non-state local institutions are characterized by decentralized governance across ethnic boundaries, they incentivize individuals to overcome collective action problems, encourage cooperative interactions, aid in the development of organizational infrastructure, and subsequently encourage the grassroots growth of ethnically-diverse civil society demanding government-provided public goods.

**Last modified:** 03-11-2021

**Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

---

## Ethnic Diversity and Public Goods Provision Across Latin America

I will control all matters in connection with this project related to data collection and publication. Curation and long-term storage of the data will be handled by the Qualitative Data Repository at Syracuse University, as detailed below. As there is only one PI, there will be no anticipated changes of responsibilities to the project.

Collected qualitative data will include semi-structured interviews from an estimated 50 to 70 local and regional government officials, high-ranking Catholic officials, Catholic laity, social movement organization leaders, and so on. Each interview is expected to last approximately forty minutes in length, on average. The themes of the interviews are not expected to put subjects at risk. However, the interviews will cover political activism. Such activism has made individuals potential targets for violent retribution in the past. Accordingly, all interview data will go through a de-identification process as detailed below.

My quantitative data consists of two originally-constructed datasets. One consists of 425 country-level time-series observations of 17 Latin American countries with data collected from the World Bank's World Development Indicators, among other sources. The other consists of more than 130,000 individual-level survey responses collected from the Latin American Public Opinion Project and other sources. Please see my Project Description for further information on data sources.

Data will be shared through the Qualitative Data Repository on publication of results in a journal article or within 12 months after project completion, whichever comes earlier.

The qualitative interviews will be collected as digital audio files via digital voice recorder and subsequently transcribed into Microsoft Word documents. All statistical data will be stored in Stata format (.dta), which can be easily exported into a variety of other highly-accessible data formats. Metadata to be made available includes interview questionnaires, instructions to respondents, informed consent forms, and de-identifying procedures, all available in Microsoft Word format and exportable as rich-text format. It also includes the Stata code used to run the statistical analyses and the Stata code used to construct the datasets, for replication purposes. These files can also be converted into rich-text format upon request. A single folder will be created for all data collected in connection with this project. Subfolders will then be created for the quantitative datasets, metadata containing qualitative interview forms, IRB protocols, and so on, and finally an encrypted file containing two subfolders: one subfolder will contain the audio files themselves, while another subfolder will contain the de-identified interview transcriptions.

Informed consent is of vital importance in conducting social research, as is the protection of the identity of respondents. This project will soon be under review with the Institutional Review Board. I have been trained to be sensitive to the ethics of research with human subjects. Prior to interviews, a written statement will be read to respondents to inform them of the nature of the research. They may then make an informed choice about whether they are willing to participate. Respondents will also be advised that, while data (including de-identified transcripts) will be made publicly available, personal and identifying information will remain confidential. I am in the process of obtaining a waiver of signed consent from the IRB in favor of verbal consent to protect participant de-identification.

A de-identification process will be developed prior to field work so that data can be de-identified as it is collected. Names will be replaced with demographic information and dates referring to specific potentially identifying events

will be replaced by general markers. Prior to public dissemination, all interview data files will be solely kept in encrypted, password-protected folders on a laptop, back-up external hard drive, and the University of Alabama's secure cloud-based server. As soon as data is collected in each region, any lists, names, or identifying information will be permanently deleted or destroyed by shredding (if physical copies exist). The digital audio files of the interviews will be permanently deleted after five years.

De-identified transcripts as well as survey data and all metadata in connection with this project will be indefinitely hosted, curated, and made freely available for research and teaching by the Qualitative Data Repository (QDR) at the Center for Qualitative and Multi-Method Inquiry at Syracuse University.[1] QDR is a third-party, National Science Foundation-funded data repository "dedicated archive for storing and sharing digital data (and accompanying documentation) generated or collected through qualitative and multi-method research in the social sciences" (<https://qdr.syr.edu/about>). QDR personnel have reviewed this plan. One-time costs for curation and storage are included in the project budget.

[1] This excludes the digital audio files which will be withheld and destroyed after five years to protect the de-identification of respondents.

Question not answered.

---