Plan Overview

A Data Management Plan created using DMPTool

Title: A Framework for Adaptive Sampling of Social Science Research Data Using the Twitter API: Understanding Social Media Communication During Crisis Events

Creator: Carl Stahmer - **ORCID:** 0000-0002-5714-3497

Affiliation: University of California, Davis

Principal Investigator: Carl Stahmer

Data Manager: Carl Stahmer

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: 18-517

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

Last modified: 02-27-2018

A Framework for Adaptive Sampling of Social Science Research Data Using the Twitter API: Understanding Social Media Communication During Crisis Events

The Project Principal Investigator, Carl Stahmer, will assume primary responsibility for all data generated during the course of this project. Whiel the primary output of the project is computer code and a Framwork for data mining, data relating to the validation of our Framework will be generated. Note that the project involves acquiring through API use and mining a large collection of Twitter data; however, the terms of the API License specifically prohibit any republication or distribution of this raw data. Secondary data that we produce will be made publicly available as open-access data sources via the California Digital Library's DASH repository service (https://www.cdlib.org/services/uc3/dash.html). DASH is the University of California, Davis' institutional repository, and there is no cost to UC Davis affiliated researchers to deposit data in DASH. Ongoing support is provided by the University and the California Digital Library. All computer code and the Framework itself will be made publicly available via the UC Davis Git Repository. This service is also supported by the University. Deposit is free of charge to UC Davis affiliated researchers.

The project will produce secondary data resulting from analysis of primary Twitter data. (Raw Twitter dats is not considered part of the project data as the Twitter API license specifically prohibits republication or sharing of this data.) Project data will indlude R data files and csv files that contain the results of various analysis of Twitter, for example, term document matricies, feature cluster matricies, document feature asssociations, etc. We will also produce computer software in the form of R scripts.

Data produced will be made available as open-access data immediately and in perpetuity. There is no plan to embargo data for any period of time.

Outputs from the project will be made available as R data files, .csv files, R code, and Microsoft Word documents. All outputs will be placed in open-access repositories where they will be freely available. There are no privacy or other licencing restirictions on the data.

Data will be stored in the UC Davis research data institutional repository (https://www.cdlib.org/services/uc3/dash.html) and the UC Davis Github repository (https://github.com/ucdavis). Both are institutionally supported and free to UC Davis affiliate researchers and both offer free and unlimited public access to data.

Question not answered.
