
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](https://orcid.org/0000-0002-5714-3497)

Affiliation: University of California, Davis (UCD)

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: 18-517

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

Last modified: 02-27-2018

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

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

Roles and responsibilities

The Data Management Plan should outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data. It must also consider changes to roles and responsibilities that will occur should a principal investigator or co-PI leave the institution.

The Project Principal Investigator, Carl Stahmer, will assume primary responsibility for all data generated during the course of this project. While the primary output of the project is computer code and a Framework 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.

Expected data

The Data Management Plan should describe the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project. It should then describe the expected types of data to be retained.

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

Period of data retention

SBE is committed to timely and rapid data distribution. However, it recognizes that types of data can vary widely and that acceptable norms also vary by scientific discipline. It is strongly committed, however, to the underlying principle of timely access, and applicants should address how this will be met in their DMP statement.

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.

Data format and dissemination

The Data Management Plan should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. Research centers and major partnerships with industry or other user communities must also address how data are to be shared and managed with partners, center members, and other major stakeholders.

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 licensing restrictions on the data.

Data storage and preservation of access

The Data Management Plan should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.

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.

Additional possible data management requirements

More stringent data management requirements may be specified in particular NSF solicitations or result from local policies and best practices at the PI's home institution. Additional requirements will be specified in the program solicitation and award conditions. Principal Investigators to be supported by such programs must discuss how they will meet these additional requirements in their Data Management Plans.

Question not answered.