

# **Data Management Plan**

## **A. Project Information**

This Data Management Plan (DMP) covers the data which will be collected for the project on ‘Organizational Responsiveness to Open Outside Input’ at the University of Massachusetts Amherst. The study projected to be conducted begin September 1, 2013 for a duration of 36 months. The study will collect non-sensitive data, most of which is already archived on free and open websites. For those data that we access via public records requests, it is the responsibility of the government agents to redact any private sensitive information before providing us with the documents. Because all of the information we would compile is non-classified and, already in the public record, we do not anticipate any confidentiality concerns.

The data collected during this study will be archived on PI Wallach’s and the UMass-Amherst department of computer science’s servers. Once the data are cleaned, organized and ready for use in scientific research, it will be made publicly available through PI Wallach’s website. We anticipate that this will be a substantially beneficial resource for researchers in the computational and social sciences.

## **B. General Data Management Plan Information**

This DMP was created on December 10, 2012 for submission to the National Science Foundation as required by NSF guidelines in the interest of securing funding for this study. The aim and purpose of this DMP is to detail and guarantee the preservation of the data collected during this study, as well as any results derived from the associated research. This DMP is intended for review by relevant NSF personnel, as well as UMASS-Amherst staff affiliated directly with this study and the collection and preservation of the associated data and research. This is the first iteration of this DMP associated with this data.

## **C. Policies**

There are no requirements stipulated by the funding or partner organizations regarding this data. Comprehensive institutional and research group guidelines specified by UMass Amherst were applied regarding the collection of this data. There are no additional requirements associated with the data being submitted.

## **D. Legal Guidelines and Requirements**

This study will only collect non-sensitive data. There are no copyright or licensing issues associated with the data being submitted.

## **E. Access, Sharing and Re-use of Data**

The researchers associated with this study are not aware of any reasons which might prohibit the sharing and re-use of the data being submitted. The researchers are not required to make this data available publicly but have elected to do so. All of the data being submitted will be made publicly available through the website of PI Wallach by September 1, 2016. There will be no additional restrictions or permissions required for accessing the data.

## **F. Data Standards and Capture**

The data will be captured using open source software tailored to scraping data from the internet and analyzed using open source machine learning software developed by the PIs. All software will be made publicly available once the PIs have used it to produce published research. The researchers are not aware of any issues regarding the effects or limitations of these formats regarding the data being submitted.

## **G. Security, Storage, Management and Back-Up of Data**

The PI's experience with, and commitment to, secure data archiving is well established. During the implementation of the study, associated research data will be physically stored on a password-protected secure server maintained by the UMass Amherst department of computer science and PI Wallach.

Research data is backed up on a daily basis. The long-term strategy for the maintenance, curation and archiving of the data will be implemented when the data and associated research are complete and ready for public distribution.

## **H. Preservation, Review and Long-Term Management of Data**

Data collected during this study will be archived with the Dataverse Network (DVN) ([thedata.org](https://thedata.org)). The data will be stored in a specific virtual archive and will be made publicly available through the DVN. As a result of this arrangement, there are no specific financial considerations of which the researchers are currently aware that might impact the long-term management of the data.