
Plan Overview

A Data Management Plan created using DMPTool

Title: Data Management Plan

Creator: Veronica Horowitz

Affiliation: University of Minnesota (twin-cities.umn.edu)

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: nsf15514

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

Last modified: 12-17-2017

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

Data Management Plan

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.

A data security plan is in place for the quantitative data used in this project. Accordingly, Co-PI Horowitz stores, manages, and analyzes the data under the guidance of PI-Uggen. Horowitz has prior experience and training with large data and the software Stata which is used for statistical modeling. Because all qualitative data is considered public data the same level of security is not required. Data will be collected by Horowitz and uploaded into NVivo for coding. Horowitz is familiar with NVivo software, having used it in other research projects, but should any software questions arise Michael Beckstrand, a qualitative software specialist and member of the support team at Liberal Arts Technology and Internet Services at UMN will be contacted for guidance. Additionally, Horowitz will train the undergrad RA in the use of NVivo software and teach the undergrad RA first the basic principles of qualitative coding, and second, more sophisticated theoretically motivated coding techniques. The Undergrad RA will help sort and organize qualitative data, and code about 15% of the qualitative data to ensure intercoder reliability.

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.

This project uses both quantitative and qualitative data. The quantitative data used is the National Corrections Reporting Program (NCRP) 2000-2014 dataset collected by the Department of Justice, Bureau of Justice statistics, containing individual case-level information on all persons incarcerated in US prisons during the time period. This data has been provided to Horowitz and Uggen from the Inter-university Consortium for Political and Social Research. In order to run multilevel models using state-year varying predictors I transform this dataset into case-year data, dramatically increasing its size, and merge the data with standard U.S. Census Bureau and Bureau of Justice Statistics sources (e.g., the annual *Statistical Abstract* and *Prisoners in the United States* series). This full dataset is about 10GB.

Three primary forms of qualitative data are used in this project. 1. Written laws, policies, statutes, and published information on commutation release, 2. Commutation applications (requested through open records laws) for at least one year (the most recent available) 3. Transcripts from commutation hearings (also collected through open records laws) and fieldnotes from attending public hearings in four states (Iowa, Louisiana, Pennsylvania, and Washington). All qualitative data will be stored in NVivo and backed up in google-drive.

Horowitz will retrieve written laws, policies, statutes, etc. primarily electronically by searching websites. Also, any relevant written materials obtained through communications with officials or nonprofits or received while attending hearings will be included. The undergrad RA will conduct a second online search to verify all important written documents have been retrieved.

All commutation applications received in the most recently available one year period have been requested by Horowitz through open records laws, received in part. These applications are varied in length (from 1 pdf page to 30) and scale (Washington has provided about 50, Pennsylvania is likely to have hundreds).

Transcripts from hearings in all four states (aiming for approximately 100 per site) have similarly been requested by Horowitz through open records laws. Two of the four research sites (Louisiana and Pennsylvania) do not transcribe their hearings. Louisiana is providing audiofiles and Pennsylvania is allowing Horowitz to videorecord hearings. Each Louisiana audiofile takes up less than 2 MB and the videofiles for one day of hearings are about 11GB in size, but both multi-media sources will be transcribed (pending funding) and kept in PDF form for most analysis. Horowitz is also attending at least two hearings in each site, and will type out and store fieldnotes in a word documents.

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.

The NCRP dataset requires periodic reauthorization for continued use. At present, the agreement between will expire on September 27th, 2019. However, should continued analysis be needed an additional one year extension will be filed. As I received this restricted dataset through the ICPSR I am not authorized to share this data, but researchers are able to request the data through ICPSR.

My qualitative data is all considered publicly available and I therefore have no ability to retain the bulk of this data since it could be requested by anyone. The exceptions to this are the fieldnotes I take from hearings along with the videofiles I am manually recording when attending hearings in Pennsylvania. Raw fieldnote data will not be disseminated at anytime. Raw video-files from Pennsylvania will not be distributed widely but will be shared if requested by interested scholars or advocates following my successful dissertation defense.

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.

I will continue to add updated documents to my compilation of written commutation laws, policies, and statutes. I will organize and store these documents in a publicly accessible drive or website. Similarly, while I cannot disseminate the NCRP dataset due to its restricted nature, my quantitative results, including visualizations of the predicted probabilities particular groups have of being released through commutation will be shared online as well.

Further, analysis of my de-identified commutation application and hearing data will be disseminated in many forms. I will promptly seek to publish my NSF-Sponsored findings in peer-reviewed journals and present these findings at academic conferences. Additionally, I will archive my dissertation online and eventually plan to write a book from my dissertation research.

Lastly, findings will be disseminated to policymakers, advocates, and clemency-focused nonprofits. Policy briefs summarizing my findings in jargon-free language will be shared with policymakers and advocates. I will create a clear and easily comprehensible summary and checklist of my findings regarding the themes that emerge in successful requests for commutation, and share this with clemency-focused nonprofits.

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.

The NCRP data will be destroyed once analysis is complete, per my data security agreement with ICPSR. The written documents, laws, and policy materials I collect from each site will be kept in google drive and updated annually. Other qualitative materials will be stored in two locations, on google drive and on an external harddrive. Google drive provides nearly unlimited storage space and is an ideal site for publicly available data storage. The external harddrive will serve as a backup in case online files are lost or destroyed. I anticipate keeping these rich qualitative data and publishing from them for at least 10 years.

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.