Data Management Plan

Proposal Number/Title: 22-2-01-2; Fueling resilience: A risk-based comparison of post-fire programs and recovery outcomes in Northern California

Principal Investigator Name and Affiliation: Dr. Miranda H. Mockrin, Northern Research Station, USDA Forest Service

I. Data Management Plan Justification

We require a data management plan as we intend to collect new data and share inputs and outcomes from wildfire modeling. We will collect two types of new data: 1) community-level recovery and mitigation actions for our study counties and 2) extent of defensible space in focal landscapes. We will share input and output data from our modeling efforts, such that researchers can replicate our results, or use our data on alternative fire modeling platforms.

II. Project Data Management

1. Data types

Participant interviews will generate several types of data on community-level recovery and mitigation actions, including participant consent forms, audio (.mp3 or .wav) recordings of interviews conducted in person, audio/video (.mp4) recordings of interviews conducted via Zoom, and written transcriptions (.docx) of audio files. Additionally, interviewees may identify publicly available planning documents related to community-level recovery and mitigation actions. Links to these documents will be appended to the final report. Using visual assessment of satellite imagery, we will generate geospatial data (.shp or .gdb) on the extent and location of defensible space, tied to buildings, for focal areas for modeling. Lastly, we will generate primary datasets as model inputs by transforming publicly available data sources (exceptions include focal area boundaries and defensible space data detailed above). We will share links to source data with other researchers through publications and will also share model input files or their sources formatted specifically for our modeling purposes. Finally we will share outputs from fire modeling runs, including individual output files (e.g., flame length, ember deposition, etc.) and aggregate results (e.g., building loss).

2. Quality Assurance

For quality assurance of interviews, we will use high quality digital sound files, either via portable recorder, or for any interviews conducted online, by recorded audio via secure Zoom teleconference software. We will employ one or more graduate students at UNT trained in qualitative research methods to transcribe and deidentify interview data. We will assess the quality of transcription by doing spot checks on the transcripts. For data analysis, we will have at least two individuals from the research team code each interview, and we will assess quality via interrater reliability statistics, a standard practice in qualitative research. For quality assurance of data collected on defensible space, we will modify Dr. Butsic's already established protocols for digitizing the spatial extent of cannabis cultivation, by dividing landscapes into grids (Butsic and Brenner 2016). The visual assessment of aerial imagery (e.g., NAIP) can lead to errors, both created by observers, and when the satellite image is obscured (e.g., clouds are present). We will assess the error rate for each student by having a sample of buildings/grids digitized by dual observers (without informing students that there is such a comparison). Students will follow standardized protocols for selecting the best images to digitize, starting with the highest-quality, most recent image, and digitizing all buildings per grid on the same images to avoid spatial shifts. The data source and timing of imagery will be collected. If it is not possible to accurately classify defensible space, we will ensure these buildings are denoted with an 'unknown' designation.

3. Data Access

Dr. Schumann will manage access to interview data. All transcriptions, notes, and audio files will be stored on password-protected project computers, using our institutions' cloud storage protocols. All interviews will be assumed to contain sensitive information (though some may not). No identifying information will

be kept with the interview notes. A separate folder will contain a document with the names of interviewees and a corresponding confidential code that will be used for all research. This document will be kept under password protection and encrypted, but in a different location on the computer than the interview transcripts themselves, in order to minimize the opportunity for participants to be identified. This identifying document will be destroyed after research is completed, and the names and codes will not be provided to anyone outside of the research team. Interview responses will not be associated with respondent names in any publications. Individuals will never be identified in any publications and precautions will be taken to ensure that individuals can never be traced back to comments. However, we will include a statement in the consent template that states: "While we will take all precautions to protect your identity, there is a risk that we could accidentally disclose information that identifies you." For the spatial data on defensible space, data entry and edits will be organized by Dr. Bustic. Each change to the data will be noted in a log, along with the reason for the change. Modeling inputs, edits, and procedures will be similarly tracked over time under Dr. Gollner's direction.

4. Storage and Backup (short-term)

Interview data (recordings, transcriptions, codes) will be managed at UNT, and will be stored on UNT's password protected OneDrive Cloud Storage and accessed on password-protected UNT computers. Spatial data on defensible space around buildings, and any other geospatial data generated/edited will be stored in excel workbooks and geodatabases, on servers through Dr. Butsic's lab at UCB. Backups of these servers include a daily incremental backup to disk. Dr. Gollner's lab at UCB uses a long-term Network Attached Storage server (43 TB) in a Raid 6 configuration to backup both raw data and model outputs. Backup of this server is accomplished through an off-site mirror and, for recent data, a redundant cloud backup. Each project partner will be responsible for periodic extra back up using external hard disks.

III. Long-Term Data Management

1.Metadata

For our geospatial data on defensible space, metadata will be documented using Metavist and we will follow the ISO 19115 metadata standard. All metadata for the spatial datasets will be Federal Geographic Data Committee compliant. Geospatial data input to ELMFIRE will be in Geospatial Tagged Image File Format (GeoTIFF) format and will be compliant with the above-mentioned standards. Additional ASCII input files will also be provided with specific input features. Each run will be separately stored within a specific folder indicating the input file changes, purpose, and where results were used. Outputs will similarly be stored in ASCII or GeoTIFF formats, depending on whether they are raw or processed outputs, respectively.

2. Data Repository

We will use the JFSP-recommended repository, Forest Service Research Data Archive (FSRDA), and transfer both defensible space and modeling input/outcomes there upon project completion. The modeling data may be downsampled to provide compiled results (e.g., probability of flame length rather than raw flame lengths) to reduce file sizes < 1TB for outputs, as is standard practice. Dr. Mockrin has a long-established working relationship with the FSRDA, and has successfully archived multiple datasets, including those from previous JFSP-funded projects.

3. Data Access

For all data archived with FSRDA, we plan to release the data set once our relevant research paper is available, or we pass the six-months post-project end date, whichever comes first. We will make our data sets "open access" rather than "monitored access" to encourage re-use. For interviews, because local government and community leaders will be discussing their roles and perspectives in overseeing specific recovery and mitigation programs, we anticipate that it will be challenging to fully de-identify interview data. Typically only limited numbers of individuals have official roles to lead such programs, and encouraging open discussion about their utility and outcomes would be constrained if such perspectives were to be made publicly available. Therefore, we will not publish interview transcripts. Instead, we will include the code lists and aggregated results from qualitative data analysis as appendices to publications that draw upon qualitative analyses, including the JFSP final report.