Data Management and Security Plan

1. What is the nature of the data?
    a. What is the form of the data?

       The data is electronic. We will be storing text responses as well as audio and video recordings

    b. Do the data contain protected health information, personal identifying information, or other sensitive information? If yes, please precisely describe what these are (see Table 1, page 3):

       Yes, the data contains personal identifying information and other sensitive information. Textual data will contain names, emails addresses, and physical addresses for payment purposes as well as the responses to pre-survey questions. Additionally, in the audio and video recordings, we will capture full face recordings and voice identifiers as the recordings will be of a zoom interview with the participant. The transcript of this recording will also be stored as text. In the interview from which we will collect data, participants will talk about their experiences with drug use (either as a user or as an outside observer) in a software engineering context.

    c. Are identifiers retained and linked to the data? Who will have access to the data? Who will have access to the identifiers?

       The names, email addresses, and physical addresses will not be retained as they are only collected for participant payment and will be collected using a separate form from the interviews (UMich Google Drive). PI Madeline Endres and RA Kaia Newman will be responsible for handling payment, and thus will be the only ones with access to these identifiers. For the recordings, however, audio voiceprints will be retained. We will destroy immediately any video recordings collected by zoom (which contain the full face identifiers). **External collaborators will not have access to names or videos.** However, we will retain the audio component of the recording. The people who will have access to the audio recordings will be researchers listed on this IRB application (Westley Weimer, Madeline Endres, Kaia Newman) as well as researchers, including external collaborators, that have been given access permission by PI Endres. **Audio recordings accessible to external collaborators will be hosted on UMich's google drive such that they can only be viewed by google users with explicit permission to do so. Before being given permission to access the audio recordings, external collaborators must go through the IRB process at their institution.** The recordings and the text transcripts will also be stored on UMich's dropbox, or MiVideo (private access). Transcripts will be created using a combination of the work done by researchers on the IRB application and Cielo24 automatic video

transcription and captioning. All of these platforms are preapproved by UMich Information Assurance for sensitive human subjects research data. Once the audio analysis is finished, audio recordings will be destroyed and we will only retain the text transcripts.

d. Are the data stripped of identifiers and the identifiers destroyed (anonymized data)? When will this take place?

   Names, emails, and physical addresses will be destroyed after participant payment is processed (no more than 2 months after the interview day). Zoom recordings collect both video and audio information. Video recordings (with full facial information) will be destroyed within 24 hours of the interview being conducted. Audio recordings will be retained throughout the analysis of the data, and thus will not be destroyed until the analysis is complete. Text transcripts will be made of the audio recordings. These will be stripped of any latent identifiers (e.g., names, workplace company name, city location).

e. Are identifiers de-linked from the data and managed by use of a code? How are the identifiers, data files and key managed and secured? Who will have access to the identifiers, data files and key?

   Identifiers (other than voice-based identifying features) will be de-linked from the data by destroying the identifiers. As these identifiers will be destroyed, no code will be used to link them to the retained data. In lieu of identifying information, all recordings and prescreening question responses will be labeled with a random number between 1 and 100 at the time of the interview. This number will be used during the analysis. External collaborators will only ever have access to data that has been de-linked from identifiers other than voice-based identifying features.

2. Where and how will the data be stored and what security measures will be used for each?

   a. Data will be stored using services that are approved by UMich for sensitive human subjects research. These services are UMich google drive, UMich Dropbox, and UMich MiVideo. For all of these services, data can only be accessed by authenticated users (via an email and password). The emails granted access to these resources will be managed by Madeline Endres and Kaia Newman. During analysis, some text snippets from the interviews might be printed on paper or note cards. Such a setup will aid the thematic analysis. When not being used for analysis, such cards will be stored in a locked file cabinet in a locked office. Once analysis is complete, any such cards will be shredded and thrown away.

b. What security measures will be used with each (password protected; encryption; locked file cabinet in locked office, 128 bit encryption, etc.)?

   All of the digital storage will be managed with password protection and authentication. The physical cards will be stored in a locked file cabinet in a locked office.

c. Who will have access to the computer/laptop/server/or files?

   Kaia Newman and Madeline Endres will both have access to the files with the direct identifying information that will be destroyed after payment is processed. For the audio files, those on the IRB will have access and additional access for external collaborators will be managed by PI Endres. For the de-identified interview transcripts, access will be managed by Madeline Endres and Kaia Newman.

3. How will data be transmitted or transported?

   a. How will electronic files be transmitted? What measures are in place for the secure transmission of data?

      Using zoom, interviews will be recorded on a laptop managed by Michigan's CSE department. **Interviews will be conducted by PI Endres and RA Newman: external collaborators will not conduct interviews.** Immediately after the interview, the audio recording will be uploaded to one of the digital UMich platforms listed above (Dropbox, MiVideo) using the UMich network (either on campus or through the VPN). Then, the video and audio recordings that were saved to the laptop will be deleted. To ensure the security of data, recordings will only be uploaded when using the UMich network.

   b. How will hardcopy files be transported?

      The only hardcopy files might be temporary notecards with snippets of de-identified textual data. These will not be transported outside of the room with the locked cabinet, and when not in use, will always be locked.

   c. How are the files and data protected while in transmission or when transported?

      Electronic data will only be transferred to one platform from the other when on the UMich Network. We will also require that all project members only access data when on a secure network (e.g., a VPN). Physical Data (note-cards for analysis) will not be transported out of the storage office with the locked cabinet.

4. When and how will data or records be deleted or destroyed?

Data records will be deleted or destroyed as follows:

- Names, Emails, and Addresses collected on google forms for payment information will be deleted from google forms once payment is processed.
- Video recordings stored to a computer will be deleted immediately after the zoom recording (moved to trash, and then trash emptied).
- Audio recordings will be deleted from the computer in a similar way as above once uploaded to UMich DropBox and/or UMich My Video. They will be deleted from those sights once the analysis is completed (6 months after the interview).

5. Will cloud-computing resources be used? (refer to UM policies at http://www.safecomputing.umich.edu/cloud/ and at http://www.safecomputing.umich.edu/google/)

    a. What is the resource and what is the privacy policy for the resource?

    Yes, the following cloud resources will be used: Umich Google Drive, Umich MiVideo, and Umich DropBox. According to the safe computing website, all of these are approved for use with Human Subjects Research.

6. Will online data collection services be used?
    a. What is the service/host? How is the survey accessed? How are data accessed by the study team? Will any non-secure services be used to access, collect, or transmit data (e.g., public portals, administrator logins, public WiFi networks, or public computers)?

    Yes, Google Forms from UMich's Google Drive will be used to collect prescreening data as well as the contact information necessary for payment, and UMich zoom will be used to record interviews. For the former, data is accessed using umich credentials and authentication. For the latter, study members and participants will be able to enter the interview using a meeting passcode that is sent through either UMich Gmail or GCal. Due to how zoom works, the recordings will be initially saved to the password-protected computer on which the subject member completed the interview. However, they will be immediately uploaded to either MiVideo and/or UMich Dropbox and destroyed on the computer. This will be done on the Umich network. No other services will be used to access, collect, or transmit data.

    b. How are data moved/transmitted from the online host to the local storage device (computer, laptop, server, thumb drive, etc.)?

    As mentioned above, due to how zoom works, the recordings will be initially saved to the password-protected computer on which the subject member

completed the interview. However, they will be immediately uploaded to either MiVideo, UMich drive, and/or UMich Dropbox and destroyed on the computer. This will be done on the Umich network (either physically or using the VPN) to protect the data. Once on the UMich services are approved for human subjects research, the recordings will not be downloaded until destroyed. They will be streamed for analysis using the Umich network. For the prescreening data collected on UMich google forms, this will never be downloaded.

c. Will the data be purged from the online host once downloaded to the local device? How and when?

   As we are primarily storing data using UMich approved online hosts (Umich dropbox and MiVideo), this will not happen. However, recordings stored on these platforms will be deleted once data analysis is complete.

d. If the data are identifiable and sensitive, are confidentiality agreements in place with outside consultants or vendors?

   No outside vendors (other than those already approved by UMich that we have mentioned earlier in this document) will be used. **We will, however, have external collaborators (e.g., from George Mason University). External collaborators will be required to go through the IRB process at their institution before being able to access audio recordings and/or textual data. Their IRB process will align with the data safety requirements in this plan and include confidentiality requirements.** After the analysis is complete, should another researcher from a different university like to be given access to the de-identified textual data, they would need to go through a data-sharing agreement process with Michigan which involves a confidentiality agreement.

7. Will any datasets be used?
   a. No datasets will be used