Get Going Data Management Plan

The data that will be generated by GetGoing will follow the following considerations:

* Ensuring privacy of the data
* Ensuring IRB compliance
* Making the data publicly available on github
* Saving the data in a format that can be used by other researchers
* Backing up the data
* Dealing with server vs cloud concerns (security)

Data collection, and management

When a user calls GetGoing, their conversation is recorded. They cannot use Get Going until they have heard the oral consent form. The speech is recorded and stored on secure servers at CMU. Even though there is no personal information collected, the GetGoing team insure that the data is truly free of personal information by reviewing the transcription of the data and eliminating any identifying items.

Data distribution

GetGoing freely distributes data that it collects to research colleagues. Upon request, a colleague is given access to the github site and permission to download. Data is key to research in transportation information, and almost any other language- related field. We see part of the results of the GetGoing project as being the data that is distributed to the community.

Similarly to what we have done on previous projects, we will be distributing the data collected through our systems in the most valuable possible ways with few restrictions to experimenters, and without compromising user privacy.

Data distribution will primarily be through the web, though in order to ensure everyone has access, we will offer a disk copy service too. From our experience with distributing data from the Let's Go Bus Information Service which include audio and text labeling, our datasets are likely to reach many 100GBs. This is reasonable for download for some groups, but not all.