SO
BC

Science
Of
Behavior
Change

Document: **Mindstrong App Data Capture and Management**
Document Status: **Final**
Date Issued: **May 2017**

MINDSTRONG APP DATA CAPTURE AND MANAGEMENT

Purpose

The purpose of this document is to define the (i) file naming convention, (ii) file retention and replication policies, (iii) file schemas, (iv) file sizes, and (v) remote configuration managment for the data captured by the Mindstrong app and analysed by the Mindstrong Digital Biomarker Platform.

Data Capture

- The Mindstrong app captures data passively which means that the user does not need to interact with the app but rather uses the phone as he or she would normally.

- The data captured by the app is stored on the device and transmitted daily to Mindstrong's AWS S3. Transmission uses secure TLS transmission. The app searches for an open wifi to begin transmission and failing that uses the cellular carrier. In this way, the Mindstrong app minimizes data plan usage.

- The sensor data captured by Mindstrong is configured to minimize battery usage. In addition, the configuration is server-side driven meaning that the settings can be adjusted dynamically for any user at any time. The Mindstrong app is consistently near the bottom of the battery consumption app list.

File Naming Convention

Files are named using the following agreed upon convention:

<study>_MINDSTRONG_<file type>_YYYYMMDDHHMMSS_<version name>.csv

where

1. <file type> is an enum with values BIOMARKER, LANGUAGE, LOCATION, SENSOR, COMMUNICATION

2. YYYYMMDDHHMMSS is the timestamp of the file generation

3. <version name> is an enum with values TEST, INTERIM, PREFINAL, FINAL

File Retention and Replication Policy

Data files are retained indefinitely in an AWS S3 bucket that is dedicated exclusively to a single study. The files are encrypted at rest using AES-256. Files are uploaded directly from the user's device to a folder identified by the participant ID.

Transmission uses TLS secure transmission. Each file is timestamped with the data of upload and is read-only. Files are replicated to a secondary AWS storage site for backup and recovery.

File Transfer Policy

The data will be transferred in four separate files using the defined file naming convention. Each file transfer will contain all cumulative data from the beginning date of the protocol. The last data transfer on completion of the protocol will be designated PREFINAL and a FINAL designation will be issued on approval by an authorized recipient.

Files will be transferred at scheduled intervals. Mindstrong provides a secure REST endpoint for file transfer.

File Schemas

The schemas for the various filetypes are found below.

The BIOMARKER data contains up to 1,035 daily markers generated from 45 keystroke behavioral and swiping/tapping patterns on the phone. Each pattern produces a daily time series of measurements from which 23 features are extracted to produce the 1,035 distinct biomarkers. Because these biomarkers measure a person's cognitive function in their day-to-day environment, they are sensitive to the effect of emotional dysregulation, sleep disturbance and other factors influencing attention, executive function, memory, processing speed and language.

The LANGUAGE data contains a daily histogram of the words used in typed communication, postings, and searches on the phone. Using natural language processing techniques we generate a time series of emotional valence measures from the histogram.

The LOCATION and SENSOR data is self explanatory.

The COMMUINICATION data contains the inbound and outbound time-stamped log of emails, sms, mms, and phone calls. The emails and phone numbers are one-way hash encrypted. This data is collected only on Android phones.

BIOMARKER

Field Name	Data Type	Description
Study Id	string	unique study id
Participant Id	string	unique participant id
OS type	string	iOS or Android
OS version	string	OS version number

Device model	string	Phone model number
Biomarker group name	string	identifies biomarker group
Biomarker subgroup name	string	identifies biomarker subgroup
Biomarker value	float	value of measured biomarker
Date	date	date of measurement

LANGUAGE

Field Name	Data Type	Description
Study Id	string	unique study id
Participant Id	string	unique participant id
OS type	string	iOS or Android
OS version	string	OS version number
Device model	string	Phone model number
Word	string	word typed
Count	integer	count aggregate over 24 hours
Date	date	date of measurement

LOCATION

Field Name	Data Type	Description
Study Id	string	unique study id
Participant Id	string	unique participant id
OS type	string	iOS or Android
OS version	string	OS version number
Device model	string	Phone model number
Latitude	float	latitude of current location
Longitude	float	longitude of current location
Altitude	float	altitude of current location
Bearing	float	horizontal direction of travel in degrees (0.0,360.0)
Speed	float	travel speed in meters per second
Accuracy	float	Radius in meters of 68% confidence circle
Timestamp	date	date and time of measurement

SENSOR

Field Name	Data Type	Description
Study Id	string	unique study id
Participant Id	string	unique participant id
OS type	string	iOS or Android
OS version	string	OS version number

Device model	string	Phone model number
Sensor type	integer	1 is accelerometer & 2 is gyroscope
First coordinate	float	first coordinate measurement
Second coordinate	float	second coordinate measurement
Third coordinate	float	third coordinate measurement
Timestamp	date	date and time of measurement

COMMUNICATION

Field Name	Data Type	Description
Study Id	string	unique study id
Participant Id	string	unique participant id
OS type	string	iOS or Android
OS version	string	OS version number
Device model	string	Phone model number
Message type	string	Email, SMS, MMS, Phone Call
Edge	0 or 1	0: inbound & 1: outbound
Node	string	One-way hash encryption of email address or phone number
Timestamp	date	date and time of measurement

FILE SIZE

Data File	Daily Size
Biomarker Raw	1-2MB
Biomarker Processed	1MB
Language	0.5MB
Location	Sampling rate dependent with 1Kb every 2 location lookups
Sensor (acc & gyro)	Sampling rate dependent but at 5Hz it is 1Kb every 2 seconds of sampling
Communication	Under 5KB

Remote Configuration Management

Control of the data that is captured and the frequency of capture is updated remotely everytime the Mindstrong app uploads data to the Mindstrong cloud. The remote configuration schema is shown below.

```
{
    entityID: entityID,
    entityType: entityType,
    gpsSensorSettings: {
        on: true,
        onDurationBackground: 30,
        offDurationBackground: 570,
        onDurationActiveInput: 30,
        offDurationActiveInput: 60
    },
    accelerometerSettings: {
        on: true,
        onDurationBackground: 10,
        offDurationBackground: 50,
        onDurationScreenOnEvent: 10,
        onDurationActiveInputEvent: 60
    },
    gyroscopeSettings: {
        on: true,
        onDurationBackground: 10,
        offDurationBackground: 50,
        onDurationScreenOnEvent: 10,
        onDurationActiveInputEvent: 60
    },
    voiceSettings: {
        on: true,
        recordOn: 120,
        recordOff: 240
    },
    uploadSettings: {
        intervalWhenWIFIDetected: 24,
        intervalWhenWIFINotDetected: 48
    },
    wordHistogramSettings: {
        on: true,
        window: 24
    },
    brainHealthTestSettings: {
        on: true,
        testBatteries: [
            {
                date: new Date(),
                tests: [
                    "FingerTapping" as Test,
                    "GoNoGo" as Test,
                    "Stroop" as Test,
                    "NBack" as Test,
                    "WorkingMemory" as Test
                ]
            }
        ]
```

```
        }
}
```