TO: Some random dude

FROM: Your loyal and devoted subject

SUBJECT: Privacy

DATE: 04/19/16

Our project necessarily involves collecting location information and thus privacy of the users should be a top priority. We need to ensure that the locations of known users can’t be determined from the data they contribute, nor can that users be identified based on the locations. There may also be privacy concerns with making WIFI access point information publically available.

**Protecting Users Individual Location Data**

Based on our other work, it seems like having some notion of personal accounts is essential for user engagement. Tying a user to the data points he collects will be necessary for ideas such as gamification and displaying to the user where they’ve mapped. Given this, it’s necessary to consider how to best provide this personalized experience to the user while not revealing any personal information in the generally available public data.

In the current application, the way a “user” is associated with the data points is by their phone’s MAC address. This is a unique identifier for the *phone*, not the user. If we hope to provide the functionality for a user to see all of the data that they have collected, then this will need to change because the MAC address is tied to the phone and thus if someone else were to get possession of the phone they would still have access to the data points collected by the previous owner.

Instead we need a user authentication system in place, whether it be our own or third party authentication such as Facebook login described by Dan in his policy memo.

**Protecting Users from Being Identified Based on “Anonymized” Location Data**

Even if each data point is anonymous, we need to be considered that the collection of data points could be used as personally identifiable information (PII). For example, given a timestamped series of GPS coordinates, it could be simple to group them into a particular source and then trace out the route of that source to determine whom it belongs to.

Until the data and userbase grows large enough, it may be impossible to completely prevent the data from identifying the users because there won’t be enough overlap between users’ routes. For this reason, we may need to explore ways to “blur” the data, or not make it available until we’re confident that it is not revealing, such as not revealing an access points until it has been recorded by some minimum number of different users.

**Potential Privacy Concerns with Making WIFI Access Point Information Broadly Available**

Collecting and publishing WIFI information is the central goal of the project, so it may be unavoidable to make it available. But it is still worth considering any potential privacy concerns with doing so. I’ve identified two cases which may be concerns:

The data would make it possible to identify the location of particular access points. Sometimes people put their names into the SSID, so this would allow someone to identify the person. Furthermore, if we were to publish the MAC address of the router, it would be possible to find the new location of a known router, even if the SSID were changed.

A hacker could also use the SSID and MAC address in order to spoof the router and trick users of that router into connecting to his router instead of the real one and then snoop on their traffic. Even though anyone could collect the same data we will be collecting, this might just facilitate the process.

For the two reasons above, I would recommend we at most only make the SSID of access points available and not the MAC address.